UNIVERSIDAD DE BUENOS AIRES

FACULTAD DE PSICOLOGÍA

ESTADÍSTICA - CÁTEDRA II

PROF. HORACIO F. ATTORRESI

GUÍA DE TRABAJOS PRÁCTICOS RESUELTA

PROF. LILIANA PACCOSI

PRÁCTICA I

EJERCITACIÓN

EJERCICIO 1

Enuncie algunas fuentes de variación que puedan afectar la calificación en un parcial de estadística y distinga las que son susceptibles de ser estudiadas como sistemáticas. Explique cómo exhibiría la información si quisiera mostrar que la asistencia a los teóricos es una fuente sistemática de variación.

Recordemos que en la variabilidad de las conductas que se estudian, distinguimos:

- las variaciones que son imprevisibles porque no están asociadas a ninguna fuente de variación sistemática, se atribuyen así a un conjunto de *fuentes fortuitas de variación*
- las variaciones que son previsibles porque están asociadas a *fuentes sistemáticas de variación*

En la calificación en un parcial de estadística, algunas fuentes de variación podrían ser: la capacidad intelectual de cada alumno, el método de estudio, el tiempo dedicado al estudio, el nivel de ansiedad en el momento del examen, el nivel de cansancio con el que el alumno llega al momento del examen, la dificultad propia del examen (en el caso que se tome un examen distinto para cada banda horaria), circunstancias individuales previas al examen, etc. Todas, salvo la última pueden ser estudiadas como sistemáticas.

La asistencia a los teóricos como fuente sistemática de variación debería exhibirse de la siguiente manera:

Calificación	Asistencia a los teóricos	
	Asiste	No asiste
10		
9		
8		
7		
6		
5		
4		
3		
2		
1		

EJERCICIO 2

Un equipo de psicopedagogos desea estudiar las habilidades adquiridas en el aprendizaje escolar de la lectoescritura mediante una prueba de rendimiento (en la escala usual de 1 a 10) y el nivel de agresividad (bajo / medio / alto) de niños provenientes de hogares con carencias económicas del Gran Buenos Aires.

Determine la población, las unidades, las características a estudiar y su nivel de medición, las variables estadísticas, su clasificación y sus valores.

Cualquier trabajo en el que se aplica la estadística se refiere a un conjunto de entidades que se conoce con el nombre de <u>*población*</u>. Llamamos población de individuos al conjunto de todos los elementos sobre los cuales se observa una o más características de Interés. En este caso la población está formada por los niños escolarizados provenientes de hogares con carencias económicas del Gran Buenos Aires.

Las *unidades* son las entidades acerca de las que se reúnen datos. El estudio que realiza el equipo de psicopedagogos se hace en niños provenientes de hogares con carencias económicas del Gran Buenos Aires, por lo tanto, las unidades están representadas por cada uno de dichos niños.

Siendo la *variable* o *característica* que se estudia, una característica de un fenómeno observable en los individuos de una población, nos encontramos con dos características a estudiar: "habilidades adquiridas en el aprendizaje de la lectoescritura" y "nivel de agresividad".

La *variable estadística* es una representación a través de números u otros símbolos de una variable. En este caso, las variables estadísticas correspondientes a las variables de estudio son: "calificación en la prueba de habilidades en lectoescritura" y "nivel de agresividad". La calificación en la prueba es una variable cuantitativa discreta, sus valores son cantidades numéricas y están representados por los números del 1 al 10 (enteros o no según la precisión adoptada) aunque la naturaleza de la característica sea continua. El "nivel de agresividad" es cuasicuantitativa y sus valores son "bajo, medio y alto", sus valores indican un rango o jerarquía. Si asignáramos a esta última variable, como valores, los números 1, 2 y 3, en lugar de bajo, medio y alto, los mismos reflejarían *sólo* los distintos grados en que se presenta esta característica, no tendría sentido decir que una persona con valor 3 es el triple de agresiva que otra con valor 1.

Al atribuir números a los objetos sólo aceptamos como válidas entre dichos números aquellas relaciones que son verificables empíricamente entre las correspondientes modalidades. Los cuatro tipos de *escalas* o *niveles de medición*, propuesto por Stevens (1946) que hemos estudiado son:

Nivel nominal: dadas dos o más modalidades, sólo podemos comprobar empíricamente si ellas son iguales o distintas, por lo tanto, entre los números atribuidos a las mismas sólo aceptamos como válida la relación igualdad-desigualdad.

Nivel ordinal: no sólo podemos comprobar si, dadas dos o más modalidades, son iguales o distintas sino también, en caso de ser distintas, cuál es la mayor entre dos de ellas. Admitiremos como válidas las relaciones de igualdad-desigualdad y orden.

Nivel de Intervalos: podemos comprobar empíricamente la igualdad-desigualdad y el orden entre dos o más modalidades, y también podemos establecer una unidad de medida y observar cuantas veces dicha unidad está contenida dentro de la diferencia entre dos modalidades cualesquiera.

Nivel de Razón: se agrega la existencia de un cero absoluto que representa la ausencia de esa característica. Esto permite interpretar razones o proporciones, es decir, podemos decir, por ejemplo, que un valor es el doble de otro.

Así, en la variable "calificación en la prueba de habilidades en lectoescritura", el nivel de medición puede considerarse de razón si la habilidad se representa por el rendimiento entendido como cantidad de respuestas correctas en la prueba. El nivel de medición de la "agresividad" es claramente ordinal por la escala adoptada.

EJERCICIO 3

Clasifique las siguientes variables e indique el nivel de medición de las escalas adoptadas.

a) Calificación de 8 individuos de un grupo según su posición en rendimiento: 1 al de mayor rendimiento, 8 al de menor rendimiento.

b) Psicodiagnóstico correspondiente a un paciente según los cuadros clínicos (neurosis, psicosis, etc.)
c) Nacionalidad de un sujeto.
d) Cantidad de palabras correctamente leídas por un disléxico en un minuto.
e) Cociente intelectual (Binet-Stern).
f) Tiempo que tarda un alumno de Psicología en concluir su carrera.
g) Temperatura en grados centígrados en Ciudad de Buenos Aires a las 0 hs.
h) Orden de mérito obtenido por un aspirante en un concurso.
i) Edad de una persona.

a) Cuasicuantitativa, nivel ordinal. En este caso, los valores asignados representan sólo un orden en el rendimiento de los 8 individuos. De dos individuos a los que se ha asignado números diferentes, puede decirse cuál de ellos presenta un mayor rendimiento, pero no cuánto mayor es el mismo.

b) Cualitativa, nivel nominal. Los valores asignados a esta variable son sólo atributos. Dados los valores correspondientes a dos pacientes, sólo podemos decir si poseen el mismo o diferente psicodiagnóstico.

c) Cualitativa, nivel nominal. La nacionalidad es un atributo del individuo e, igual que en el punto anterior, de dos sujetos sólo diremos si poseen la misma o diferente nacionalidad.

d) Cuantitativa discreta, nivel de razón. El cero de la variable es absoluto, indica que una persona disléxica no pudo leer correctamente al menos una palabra en un minuto. Es una variable discreta ya que no es posible que la persona lea, por ejemplo, dos palabras y media.

e) Cuantitativa continua, nivel intervalar o Cuasicuantitativa, nivel ordinal (es discutible). Una persona que posee un IQ de 100, tiene una inteligencia más cercana a alguien con un IQ de 110 que otra con un IQ de 60; parece ser una escala de intervalos, pero es difícil establecer que la escala realmente posea intervalos *iguales* entre las unidades adyacentes. Sin embargo, muchos investigadores consideran tales variables como si fuesen medidas en escalas de intervalos, en particular cuando el instrumento de medición es bastante estándar, como es el caso del WAIS.

f) Cuantitativa continua, nivel de razón (el cero de la escala no tiene por qué ser un valor posible para la variable).

g) Cuantitativa continua, nivel intervalar (el cero de la escala no es absoluto, no representa realmente ausencia de temperatura).

h) Cuasicuantitativa, nivel ordinal (los valores asignados a cada aspirante indican un orden o jerarquía).

i) Cuantitativa discreta (pues se considera en años o meses enteros), nivel de razón. El cero es una convención de la definición de “edad” pero no una convención de la escala. En nuestra cultura la edad se define como cantidad de años enteros (o meses para los pequeños) que transcurren a partir del nacimiento.

EJERCICIO 4

Una Consultora de Recursos Humanos está interesada en estudiar el clima laboral de una organización. Con este fin se efectúa una evaluación objetiva al contar la cantidad de “quejas” diarias que brindan 30 de sus 150 empleados.

a) ¿Cuál es la característica objeto de estudio y cuál es la variable estadística con la que se la operacionaliza?

b) Clasifique y determine el nivel de medición de la variable.
c) Determine la población de individuos y muestra de individuos
d) Determine la población de observaciones y la muestra de observaciones.

a) La característica: clima laboral percibido por los empleados. La variable estadística: La manifestación observable que da cuenta del clima laboral percibido por los empleados es la “cantidad de quejas diarias de los empleados”.

b) Clasificación y nivel de medición de la variable: Cuantitativa discreta, la cantidad de quejas es un número entero no negativo. El nivel de medición es de cociente o razón, pues el cero en la escala indica que el empleado no ha realizado ninguna queja.

c) La población de individuos está constituida por el conjunto de empleados de la organización (que en este caso son 150). La muestra de individuos está constituida por las 30 empleados de la organización seleccionados.

d) Siendo la población de observaciones el conjunto de todos los valores que puede tomar la variable estadística sobre la población de individuos, la misma está constituida por el conjunto de números que representan a la cantidad de quejas diarias de los 150 empleados de la organización. La muestra de observaciones es el conjunto de valores que toma la variable estadística pero, ahora, sobre la muestra de individuos; por lo tanto, está constituida por el conjunto de números que representan a la cantidad de quejas diarias de los 30 empleados de la organización seleccionados.

EJERCICIO 5

El Centro de estudiantes de la Facultad de Psicología de la UBA está interesado en conocer el grado de acuerdo de distintos actores de dicha facultad, respecto de la inclusión de la materia Matemática en el CBC de la carrera. Para ello al finalizar el primer cuatrimestre de 2005, tomó una muestra aleatoria de 1500 estudiantes y 400 docentes de la carrera. A continuación se presentan los resultados (no reales):

Grado de acuerdo	*Alumnos*	*Docentes*
De acuerdo	20%	40%
Indiferente	10%	30%
En desacuerdo	70%	30%
Total	100%	100%

a) Determine quiénes constituyen la población y quiénes la muestra de individuos.
b) Determine la población y la muestra de observaciones
c) Mencione la variable, clasifíquela e indique la escala de medición.
d) Indique la posible fuente sistemática que provoque variaciones previsibles en el grado de acuerdo con la inclusión de la materia Matemática en el CBC.

El estudio se lleva a cabo sobre **estudiantes** y **docentes** de la Facultad de Psicología de la UBA, por lo tanto:

a) Hay dos poblaciones de individuos: una está constituida por el conjunto de estudiantes de la carrera de Psicología de la UBA, y la otra por el conjunto de docentes de la carrera de Psicología de la UBA. Hay dos muestras de individuos: una está formada por los 1500 estudiantes seleccionados de la carrera de Psicología de la UBA y la otra por los 400 docentes seleccionados de la misma carrera.

Dicho estudio recoge la opinión de ambos actores respecto del **grado de acuerdo** con la inclusión de la materia Matemática en el CBC de la carrera; en correspondencia con las dos poblaciones y muestras de individuos:

b) Hay dos poblaciones de observaciones: una está formada por el conjunto de valores que representan el grado de acuerdo con la inclusión de Matemática en el CBC para la carrera de Psicología entre los alumnos y la otra por el conjunto valores que representan el grado de acuerdo con la inclusión de Matemática en el CBC para la carrera de Psicología entre los docentes. Hay dos muestras de observaciones: una está formada por el conjunto de valores que representan el grado de acuerdo con la inclusión de Matemática en el CBC para la carrera de Psicología entre los 1500 alumnos seleccionados y la otra, por el conjunto valores que representan el grado de acuerdo con la inclusión de Matemática en el CBC para la carrera de Psicología entre los 400 docentes seleccionados.

c) Variable: Grado de acuerdo con la inclusión de Matemática en el CBC para la carrera de Psicología. Es una variable Cuasicuantitativa, sus valores indican un orden o jerarquía. Nivel de medición Ordinal.

d) Fuente sistemática de variación: tipo de actor - alumno o docente -. Entre los alumnos se observa un porcentaje mayor que está en desacuerdo con la inclusión de la materia Matemática en la carrera, mientras que entre los docentes se observa un mayor grado de acuerdo con la inclusión de la misma.

EJERCICIO 6

Para comparar la eficiencia de un método de autoinstrucción con el tradicional en la enseñanza del tema “Prueba de hipótesis” (Estadística Inferencial) se han elegido al azar 40 alumnos de la cátedra I de Estadística del primer cuatrimestre de 2009. Fueron asignados aleatoriamente 20 a cada modalidad de enseñanza. Los resultados de su posterior evaluación son:

Calificación	Número de alumnos	
	Método tradicional	Método de autoinstrucción
3	1	
4	5	1
5	7	1
6	4	2
7	3	7
8		5
9		3
10		1

a) Lea la información de la tabla.
b) Determine quiénes constituyen la población y quiénes la muestra de individuos y de observaciones.
c) Mencione la variable estadística y clasifíquela.
d) ¿Observa alguna tendencia en las observaciones que den cuenta de la existencia de una fuente sistemática de variación? Si su respuesta es afirmativa indique cuál.
e) ¿Todas las calificaciones obtenidas bajo el método de autoinstrucción fueron mayores que las obtenidas bajo el método tradicional? ¿A qué atribuye esto?
f) Mencione posibles fuentes fortuitas de variación que afecten la calificación.
g) Explique, a partir de d) y e) por qué se acudió a una serie de observaciones en cada grupo para compararlos en lugar de usar un solo representante de cada grupo.

a) Hay un alumno instruido con el método tradicional que obtuvo una calificación de 3 puntos, cinco alumnos instruidos con el método tradicional obtuvieron una calificación de 4, siete alumnos obtuvieron 7 con el método de autoinstrucción, etc.

b) La población de individuos: está integrada por la totalidad de los alumnos de la Cátedra I de Estadística que cursaron la materia en el primer cuatrimestre de 2009.
Muestra de individuos: hay dos muestras de individuos, cada una de 20 alumnos, que son los que participaron de la experiencia.
Población de observaciones: hay dos poblaciones hipotéticas de observaciones, una está formada por las calificaciones que habrían obtenido los alumnos de la Cátedra I de Estadística del primer cuatrimestre de 2009 en la evaluación sobre prueba de hipótesis, si éstos hubieran sido instruidos por el método tradicional, y la otra población hipotética la integran las calificaciones que habrían obtenido bajo el método de autoinstrucción dichos alumnos.
Muestra de observaciones: hay dos muestras de observaciones, una, las calificaciones bajo el método tradicional obtenidas por los 20 alumnos de este grupo, y la otra, las calificaciones bajo el método de autoinstrucción obtenidas por los otros 20 alumnos.

c) La variable estadística es la calificación obtenida por los alumnos en la evaluación acerca del tema “prueba de hipótesis”. Es una variable cuantitativa discreta.

d) Sí. Las calificaciones tienden a ser mayores para el grupo correspondiente al método de autoinstrucción. Aparentemente el método de instrucción es una fuente sistemática de variación para la calificación. Decimos “aparentemente” porque para llegar a la conclusión utilizamos la simple inspección ocular (más adelante veremos métodos adecuados de decisión estadística).

e) No, debido a las fuentes fortuitas de variación.

f) Capacidad individual, preferencia del sujeto a recibir un determinado tipo de instrucción, el docente a cargo de cada uno de los métodos, etc.

g) Si no existieran fuentes fortuitas de variación y toda la variabilidad en la calificación se debiera al método, entonces todas las calificaciones coincidirían dentro del mismo método y bastaría tomar una observación de cada uno para compararlos. Pero al existir otras fuentes de variación es necesario hacer una “serie de observaciones” para que el efecto de las fuentes fortuitas se vea compensado en el conjunto y deje ver la tendencia en el comportamiento de la variable. Por ejemplo: si se hubiera tomado una sola observación de cada grupo podría haber ocurrido que se eligiera el alumno que sacó 7 con el método tradicional y el que sacó 4 con el de autoinstrucción; no se sabría entonces si estas calificaciones son debidas a que el método tradicional es mejor o a que el azar hizo que el alumno más inteligente recibiera el método tradicional. Pero cuando a cada método se asigna al azar un conjunto de individuos, es poco probable que todos los más inteligentes queden en uno de ellos y el efecto de la inteligencia se verá compensado en dicho conjunto sin ocultar el efecto de la fuente sistemática.

EJERCICIO 7

Los datos de este ejercicio fueron obtenidos de un largo estudio por Mr. Joseph Raffaelle, de la universidad de Pittsburgh, para analizar la comprensión lectora (C) y la rapidez lectora (R), usando los puntajes obtenidos en los subtests del IOWA TEST OF BASIC SKILLS.

Después de seleccionar aleatoriamente 30 estudiantes y dividirlos al azar en 6 submuestras de tamaño 5, los grupos fueron asignados aleatoriamente a dos tratamientos (clases abreviadas y clases no abreviadas) y tres maestros. Los datos son:

	Clases abreviadas		Clases no abreviadas	
	R	C	R	C
Maestro 1	10	21	9	14
	12	22	8	15
	9	19	11	16
	10	21	9	17
	14	23	9	17
Maestro 2	11	23	11	15
	14	27	12	18
	13	24	10	16
	15	26	9	17
	14	24	9	18
Maestro 3	8	17	9	22
	7	15	8	18
	10	18	10	17
	8	17	9	19
	7	19	8	19

a) Indique cuáles son las posibles fuentes sistemáticas de variación.
b) Mencione las características a estudiar, su nivel de medición, las variables estadísticas y su clasificación.
c) Mencione posibles fuentes fortuitas de variación.

a) Las posibles fuentes sistemáticas de variación son la modalidad de la clase y el maestro.

b) Las características por estudiar: “rapidez lectora” y “comprensión lectora”.
El nivel de medición: de razón para la primera y discutible para la segunda, puede ser ordinal o intervalar (el tratamiento estadístico posterior que de ella hace el autor corresponde al nivel intervalar).
Las variables estadísticas: “calificación en rapidez lectora” – cuantitativa discreta, aunque la característica sea continua – y “calificación en comprensión lectora” – cuasicuantitativa o cuantitativa discreta según el nivel de medición adoptado - .

c) Posibles fuentes fortuitas de variación: la capacidad individual, el entrenamiento extra-escolar, el nivel cultural de los estudiantes, etc.

EJERCICIO 8 (Para reflexionar y discutir en conjunto)

En Attorresi et al. (2008) se presenta un cuestionario para medir la actitud altruista. Ésta consiste de 18 ítems con dos opciones: una corresponde a una actitud altruista y la otra no. El puntaje que se asigna al individuo corresponde a la cantidad de respuestas que denotan una actitud altruista.
Suponiendo que Enrique, Luis, y Matías hayan obtenido 12 puntos, 16 puntos y 8 puntos respectivamente:

a) Responda justificando las respuestas. ¿Afirmaría que...
    a1) ... la actitud altruista de Enrique difiere de la de Luis en la misma medida que difiere de la actitud altruista de Matías?
    a2) ... Luis es el doble de altruista que Matías?
b) Mencione la característica objeto de estudio, su nivel de medición, el instrumento de medición y la variable estadística.

a) 1) y 2) No parece realista cuantificar la actitud altruista con tal precisión que permita dar una métrica de la diferencia en altruismo entre uno y otro individuo.

b) La característica es la actitud altruista.
Nivel de medición: más propiamente ordinal, pues tiene sentido pensar que una persona es más o menos altruista que otra, inclusive podría decirse "mucho más" o "mucho menos" altruista; es decir, podría pensarse en la proximidad o distancia entre las personas en cuanto a su actitud altruista en términos "ordinales" pero cuantificarla con la precisión que supone el nivel intervalar parece muy pretencioso. Parece más razonable interpretar los puntajes como rangos que como cantidades.
El instrumento de medición: es el cuestionario.
La variable estadística: es la "cantidad de respuestas que denotan la actitud altruista". Obsérvese que si se tomara esta variable como objeto de estudio en sí misma sin remitirnos al significado en términos de la característica que representa, correspondería al nivel de razón, ya que es una variable intrínsecamente cuantitativa (cantidad de respuestas...) y el cero no es arbitrario, significa que el individuo no dio ninguna respuesta que denotara una actitud altruista. Si a la variable se le diera un tratamiento estadístico correspondiente al nivel de razón habría que tener en cuenta que todo lo que se concluye se referirá al puntaje y podría interpretarse en términos de actitud altruista sólo en tanto y en cuanto se considere que dicho puntaje refleja con suficiente precisión a la característica en cuestión (actitud altruista).

EJERCICIO 9

Al registrarse el puntaje de 5 sujetos en una prueba de razonamiento lógico y el tiempo empleado para resolverla se obtuvieron las siguientes observaciones apareadas:

Puntaje X (sobre un total de 20):	15	13	20	15	18
Tiempo Y (en minutos):	60	55	68	50	65

Verifique si son iguales o no las siguientes expresiones. En caso de ser iguales decidir si lo son para cualquier conjunto de observaciones.

a)	$\sum_{i=1}^{5} x_i^2$	y	$\left(\sum_{i=1}^{5} x_i\right)^2$
b)	$\sum_{i=1}^{5} 2y_i$	y	$2\sum_{i=1}^{5} y_i$
c)	$\sum_{i=1}^{5} x_i . y_i$	y	$\sum_{i=1}^{5} x_i . \sum_{i=1}^{5} y_i$

La letra griega mayúscula sigma, Σ, indica la operación de sumatoria. La expresión algebraica para la sumatoria es:

$$\sum_{i=1}^{n} x_i$$

que se lee como "suma de todos los valores que toma la variable *x*, desde $i = 1$ a n".

a) La expresión $\sum_{i=1}^{5} x_i^2$ indica la suma de los cuadrados de los datos, es decir que, primero se debe elevar al cuadrado cada puntuación X y luego sumar los valores así obtenidos.

En cambio, la expresión $\left(\sum_{i=1}^{5} x_i\right)^2$ indica el cuadrado de la suma de los cinco valores, es decir que, primero debemos sumar los cinco valores y luego elevar al cuadrado la suma resultante.

$$\sum_{i=1}^{5} x_i^2 = 15^2 + 13^2 + 20^2 + 15^2 + 18^2 = 1343$$

$$\sum_{i=1}^{5} x_i^2 = (15 + 13 + 20 + 15 + 18)^2 = 6561$$

Son expresiones diferentes, ya que la potenciación no es distributiva respecto de la suma.

b) La expresión $\sum_{i=1}^{5} 2y_i$ indica que cada puntuación Y se multiplica por dos y luego se suman los productos parciales.

En la expresión $2\sum_{i=1}^{5} y_i$, la suma de los cinco valores se multiplica por 2.

$$\sum_{i=1}^{5} 2y_i = 2·60 + 2·55 + 2·68 + 2·50 + 2·65 = 596$$

$$2\sum_{i=1}^{5} y_i = 2·(60 + 55 + 68 + 50 + 65) = 596$$

Ambas expresiones son iguales. Esto se verifica por la propiedad distributiva de la multiplicación respecto de la suma.

c) Esta expresión $\sum_{i=1}^{5} x_i.y_i$ indica que debemos multiplicar los valores con igual índice, es decir, el primer valor del puntaje X con el primer valor del puntaje Y, el segundo valor de X con el segundo valor de Y, y así hasta el último, y luego sumar todos los productos así obtenidos.

En cambio, $\sum_{i=1}^{5} x_i.\sum_{i=1}^{5} y_i$ indica que debemos multiplicar el resultado de la suma del primer grupo de valores X por el resultado de la suma del segundo grupo de valores Y.

$$\sum_{i=1}^{5} x_i.y_i = 15\cdot 60 + 13\cdot 55 + 20\cdot 68 + 15\cdot 50 + 18\cdot 65 = 4895$$

$$\sum_{i=1}^{5} x_i.\sum_{i=1}^{5} y_i = (15 + 13 + 20 + 15 + 18)\cdot(60 + 55 + 68 + 50 + 65) = 81\cdot 298 = 24.138$$

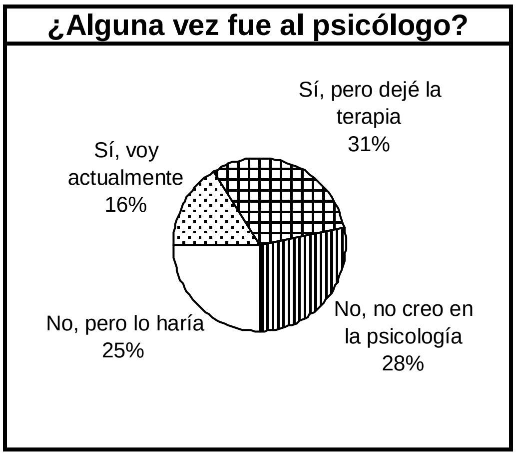
Las expresiones obtenidas resultan distintas.

PRÁCTICA II

EJERCITACIÓN

EJERCICIO 1

El siguiente diagrama circular muestra los resultados de una encuesta de opinión llevada a cabo sobre 30.646 personas de Ciudad de Buenos Aires y Gran Buenos Aires, publicada en el diario Clarín el 26 de octubre de 2005.



a) Presente la información en una tabla donde se mencione la variable, se muestren sus valores, los ángulos de cada sector y la cantidad de personas que sustenta cada opinión.
b) Mencione el nivel de medición de la variable.
c) Indique si la escala está bien diseñada en términos de las propiedades que deben caracterizarla. Fundamente la respuesta.

a)
- La variable presente en la encuesta es “Hábito o Disposición de los habitantes de Ciudad de Buenos Aires y Gran Buenos Aires para frecuentar al psicólogo”. Es una variable cualitativa, ya que sus valores sólo expresan atributos, son nombres o categorías.

- La cantidad de personas que sustenta cada opinión se obtiene calculando el porcentaje respectivo directamente con una calculadora científica, construyendo una regla de tres simple, o expresando dicho porcentaje en su notación decimal.

Ejemplo:

100 % ———————— 30.646 personas
16 % ———————— x personas

$$x = \frac{16\ \% \cdot 30.646}{100\ \%} = 4903{,}36$$

Es decir, **4903 personas** sustentan la opinión “**Sí, voy actualmente**”.

Se obtiene el mismo resultado multiplicando $0{,}16 \cdot 30.646 = 4903{,}36$

Análogamente se calculan los otros porcentajes:

$0{,}31 \cdot 30.646 = 9500{,}26$

9500 personas sustentan la opinión “**Sí, pero dejé la terapia**”.

$0{,}28 \cdot 30.646 = 8580{,}88$

8581 personas sustentan la opinión “**No, no creo en la psicología**”.

$0{,}25 \cdot 30.646 = 7661{,}5$

7662 personas sustentan la opinión "**No, pero lo haría**"

- El ángulo de cada sector se obtiene también mediante una regla de tres simple, ya que la medida de cada ángulo es directamente proporcional al porcentaje de personas que le corresponde a cada sector.

$$\begin{array}{rcl} 100\ \% & \text{———————} & 360° \\ 16\ \% & \text{———————} & x° \end{array}$$

$$x = \frac{16\ \% \cdot 360°}{100\ \%} = 57{,}6° = 57°36'$$

o también,

$$x = 0{,}16 \cdot 360° = 57{,}6° = 57°36'$$

Análogamente:

$$x = 0{,}31 \cdot 360° = 111{,}6° = 111°36'$$

$$x = 0{,}28 \cdot 360° = 100{,}8° = 100°48'$$

$$x = 0{,}25 \cdot 360° = 90°00'$$

La siguiente tabla muestra los resultados calculados anteriormente:

Variable: "Hábito o Disposición de los habitantes de Ciudad de Buenos Aires y Gran Buenos Aires para frecuentar al psicólogo".	Ángulos	Cantidad de personas
Sí; voy actualmente.	57º 36'	4903
Sí pero dejé la terapia.	111º 36'	9500
No; no creo en la Psicología.	100º 48'	8581
No pero lo haría.	90º 00'	7662

b) El nivel de medición es nominal, pues si conocemos las respuestas dadas por dos o más personas sólo podemos comprobar si pertenecen a la misma o a distinta categoría; la única relación que aceptamos como válida entre los valores que toma la variable es la relación de igualdad-desigualdad.

c) Las categorías cumplen la condición de ser mutuamente excluyentes y, con respecto a la información de hecho obtenida y en aras de una síntesis, también pueden considerarse colectivamente exhaustivas. Sin embargo, a priori, también quizás se hayan considerado las categorías NS/NC (No Sabe/No Contesta) y Otras; esta última suele incluirse para lograr la exhaustividad ante la posibilidad o el hecho de que se presenten respuestas con baja frecuencia; por ejemplo: "No iría porque no soy perseverante para las terapias".

EJERCICIO 2

El siguiente cuadro resume los resultados obtenidos en una investigación de Inhelder y Vinh-Bang (Piaget & Inhelder, 1970, pág.16), en la que se estudió la conservación de la sustancia, peso y volumen de acuerdo con cómo son percibidos por niños de 5 a 11 años. En la siguiente tabla se muestran los resultados expresados **en porcentajes**:

		Edad						
		5	6	7	8	9	10	11
Sustancia	No conservación	84	68	**64**	**24**	12	-	-
	Intermedio	0	16	**4**	**4**	4	-	-
	Conservación	16	16	**32**	**72**	84	-	-
Peso	No conservación	100	84	76	40	16	16	0
	Intermedio	0	4	0	8	12	8	4
	Conservación	0	12	24	52	72	76	96
Volumen	No conservación	100	100	88	44	56	**24**	16
	Intermedio	0	0	0	28	12	**20**	4
	Conservación	0	0	12	28	32	**56**	80

a) Mencione la/s variable/s y la fuente sistemática de variación.
b) Calcule los ángulos de un diagrama circular correspondiente a la distribución de frecuencias de la variable nivel de conservación del volumen en niños de 10 años.
c) Indique el gráfico adecuado para comparar a los niños de 7 y 8 años en la percepción del nivel de conservación de las sustancias.

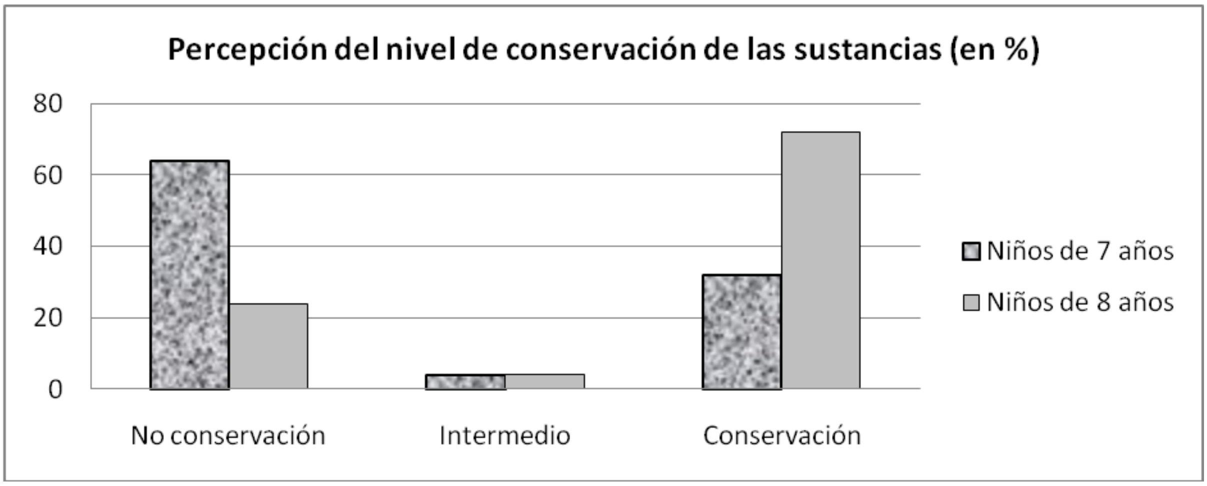
a) El cuadro contiene la información acerca de tres variables: “Percepción del nivel de conservación de sustancia”, “Percepción del nivel de conservación de peso” y “Percepción del nivel de conservación de volumen”. La fuente sistemática de variación es la “edad” de los niños.

b) Se deben calcular los ángulos correspondientes a los porcentajes 24%, 20% y 56%:

$$0{,}24 \cdot 360° = 86{,}4° = 86°24'$$
$$0{,}20 \cdot 360° = 86{,}4° = 72°$$
$$0{,}56 \cdot 360° = 201{,}6° = 201°36'$$

Los ángulos son: 86° 24’ (No conservación), 72° (Intermedio), 201° 36’ (Conservación)

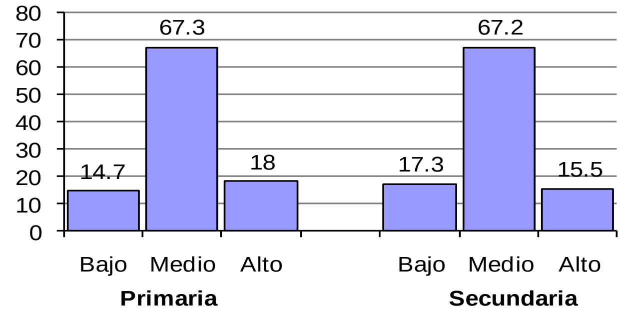
c) Una posibilidad es el diagrama de rectángulos adyacentes. Este tipo de gráfico permite comparar dos o más conjuntos de datos diferentes; en este caso el conjunto de valores correspondientes a los niños de 7 años con el conjunto de valores de los niños de 8 años.
A continuación, se muestra, a modo ilustrativo, el diagrama para ambas distribuciones. Las frecuencias porcentuales a utilizar para cada grupo son 64, 4 y 32 para los niños de 7 años, y 24, 4 y 72, para los niños de 8 años.


Percepción del nivel de conservación de las sustancias (en %)

EJERCICIO 3

A continuación se exhibe una parte de los resultados de una encuesta de opinión sobre el nivel de la enseñanza primaria y secundaria. (Clarín, 27 / 02 /1994).

OPINIÓN SOBRE EL NIVEL DE LA ENSEÑANZA
según alumnos y ex alumnos de 14 a 24 años. **(En %)**



Ficha Técnica

- Tipo de encuesta: cuestionario estructurado con preguntas cerradas y abiertas. Las preguntas de este estudio forman parte de un cuestionario más amplio que abarca temas sociales y económicos.
- Tamaño de la muestra: 1019 jóvenes (14 a 24 años) y 1002 adultos (25 años y más)
- Tipo de muestra: probabilística y por cuotas según sexo, edad y ocupación.
- Área geográfica: Ciudad de Buenos Aires y Gran Buenos Aires
- Trabajo de campo: agosto 1992

Fuente: Demoskopia.

a) ¿Cuántas distribuciones están representadas? Mencione las variables correspondientes, indique el nivel de medición de la escala utilizada y sus valores.
b) Lea la ficha técnica. Vuelque los datos correspondientes a las opiniones sobre la primaria en tablas de distribución de frecuencias absolutas, relativas y porcentuales con sus correspondientes acumuladas. Si lo cree necesario para su ejercitación repítalo para la otra distribución.
c) ¿Cuántos encuestados consideran que el nivel de enseñanza secundaria es, a lo sumo, medio?
d) ¿Qué porcentaje considera que la enseñanza primaria es, por lo menos, medio?

a) Están representadas en el mismo diagrama dos distribuciones correspondientes a las variables “opinión de los jóvenes de 14 a 24 años respecto del nivel de la enseñanza primaria” y “opinión de los jóvenes de 14 a 24 años respecto del nivel de la enseñanza secundaria”. Se trata de variables cuasicuantitativas, el nivel de medición es ordinal, ya que sus valores indican un rango o jerarquía; dichos valores son bajo, medio y alto.

b) La **frecuencia absoluta** es la cantidad de veces que cada valor de la variable aparece en un conjunto de datos.

Para obtenerla, calculamos los porcentajes correspondientes a cada valor de la opinión de los 1019 jóvenes, sobre el nivel de la **enseñanza primaria**:

Nivel Bajo: 14,7% de 1019, es decir, $0{,}147 \cdot 1019 = 149{,}793$
150 personas

Nivel Medio: 67,3% de 1019, es decir, $0{,}673 \cdot 1019 = 685{,}787$
686 personas

Nivel Alto: 18% de 1019, es decir, $0{,}18 \cdot 1019 = 183{,}42$
183 personas

Una forma de comprobar que no se cometieron errores es comprobar que la suma de todos los valores obtenidos sea igual a ***n*** (tamaño de la muestra).

150 + 686 + 183 = 1019

La **frecuencia absoluta acumulada** de un valor es el número de veces que se repite en la muestra ese valor o cualquier otro valor inferior. Para obtenerla, sumamos para cada valor su frecuencia absoluta más la absoluta acumulada del valor anterior. La frecuencia acumulada del valor mayor es igual a ***n***.

Nivel Bajo: 150
Nivel Medio: 686 + 150 = 836
Nivel Alto: 183 + 836 = 1019

Para obtener la **frecuencia relativa** dividimos cada frecuencia absoluta por ***n***. La suma de todas las frecuencias relativas es igual a 1. A veces esta columna no suma aritméticamente 1, debido al redondeo que se hace en su cálculo, no es raro encontrar que esta columna da valores iguales a 0,98 o 0,99.

Nivel Bajo: $\frac{150}{1019} \approx 0{,}147$

Nivel Medio: $\frac{686}{1019} \approx 0{,}673$

Nivel Alto: $\frac{183}{1019} \approx 0{,}180$

0,147 + 0,673 + 0,18 = 1

Para calcular la **frecuencia relativa acumulada** dividimos cada frecuencia absoluta acumulada por ***n***. También podemos sumar para cada valor su frecuencia relativa más la relativa acumulada del valor anterior. La frecuencia relativa del valor mayor es igual a 1.

Nivel Bajo: $\frac{150}{1019} \approx 0{,}147$

Nivel Medio: $\frac{836}{1019} \approx 0{,}820$ o también 0,147 + 0,673 = 0,820

Nivel Alto: $\frac{1019}{1019} = 1$ o también 0,820 + 0,18 = 1

Para obtener las **frecuencias porcentuales**, multiplicamos cada frecuencia relativa y relativa acumulada por 100:

Nivel Bajo: $0{,}147 \cdot 100 = 14{,}7$

Nivel Medio: $0{,}673 \cdot 100 = 67{,}3$

Nivel Alto: $0{,}18 \cdot 100 = 18$

La suma de las frecuencias porcentuales es igual a 100 (los valores obtenidos son los observados, como resultado de la encuesta, en el gráfico).

Calculamos las porcentuales acumuladas:

Nivel Bajo: $0{,}147 \cdot 100 = 14{,}7$

Nivel Medio: $0{,}82 \cdot 100 = 82$

Nivel Alto: $1 \cdot 100 = 100$

A continuación se realizan los cálculos correspondientes a cada valor de la opinión de los 1019 jóvenes, sobre el nivel de la **enseñanza secundaria**:

Frecuencias absolutas:

Nivel Bajo: 17,3% de 1019, es decir, $0{,}173 \cdot 1019 = 176{,}287$
176 personas

Nivel Medio: 67,2% de 1019, es decir, $0{,}672 \cdot 1019 = 684{,}768$
685 personas

Nivel Alto: 15,5% de 1019, es decir, $0{,}155 \cdot 1019 = 157{,}945$
158 personas

Frecuencias absolutas acumuladas:

Nivel Bajo: 176
Nivel Medio: 685 + 176 = 861
Nivel Alto: 158 + 861 = 1019

Frecuencias relativas:

Nivel Bajo: $\frac{176}{1019} \approx 0{,}173$

Nivel Medio: $\frac{685}{1019} \approx 0{,}672$

Nivel Alto: $\frac{158}{1019} \approx 0{,}155$

0,173 + 0,672 + 0,155 = 1

Frecuencias relativas acumuladas:

Nivel Bajo: $\frac{176}{1019} \approx 0{,}173$

Nivel Medio: $\frac{861}{1019} \approx 0{,}845$ o también 0,672 + 0,173 = 0,845

Nivel Alto: $\frac{1019}{1019} = 1$ o también 0,155 + 0,845 = 1

Frecuencias porcentuales:

Nivel Bajo: $0{,}173 \cdot 100 = 17{,}3$

Nivel Medio: $0{,}672 \cdot 100 = 67{,}2$

Nivel Alto: $0{,}155 \cdot 100 = 15{,}5$

Frecuencias porcentuales acumuladas:

Nivel Bajo: $0{,}173 \cdot 100 = 17{,}3$

Nivel Medio: $0{,}845 \cdot 100 = 84{,}5$

Nivel Alto: $1 \cdot 100 = 100$

La siguiente tabla muestra los cálculos efectuados anteriormente.

Opinión	Enseñanza Primaria						Enseñanza Secundaria					
	n_i	n_a	p_i	p_a	P_i%	P_a%	n_i	n_a	p_i	p_a	P_i%	P_a%
Alto	183	1019	0,18	1	18	100	158	1019	0,155	1	15,5	100
Medio	686	836	0,673	0,82	67,3	82	685	861	0,672	0,845	67,2	84,5
Bajo	150	150	0,147	0,147	14,7	14,7	176	176	0,173	0,173	17,3	17,3
	1019		1		100		1019		1		100	

c) La cantidad de encuestados que consideran que el nivel de enseñanza secundaria es **a lo sumo** medio, está determinada por aquellos encuestados que consideran que el nivel de enseñanza es bajo o medio, es decir, debemos sumar 176 (bajo) más 685 (medio), lo que da un total de **861 personas**. También se obtiene el mismo resultado, sumando los porcentajes correspondientes a dichos valores (17,3 y 67,2) y calculando este porcentaje sobre los 1019 jóvenes encuestados:

17,3% + 67,2% = 84,5%
0,845 · 1019 = 861,055, es decir, 861 personas.

d) Las personas que consideran que el nivel de enseñanza primaria es **por lo menos** medio, significa que consideran que es medio o alto, por lo tanto, debemos sumar los porcentajes 67,3 (medio) y 18 (alto). El porcentaje buscado es **85,3%**.

EJERCICIO 4

Estos son los resultados (no reales) de una encuesta efectuada a graduados de nuestra facultad que respondieron cuál ha sido la **asignatura más compleja del primer año de la carrera**. Observe la siguiente distribución de **frecuencias relativas** e identifique los ERRORES cometidos en su confección.

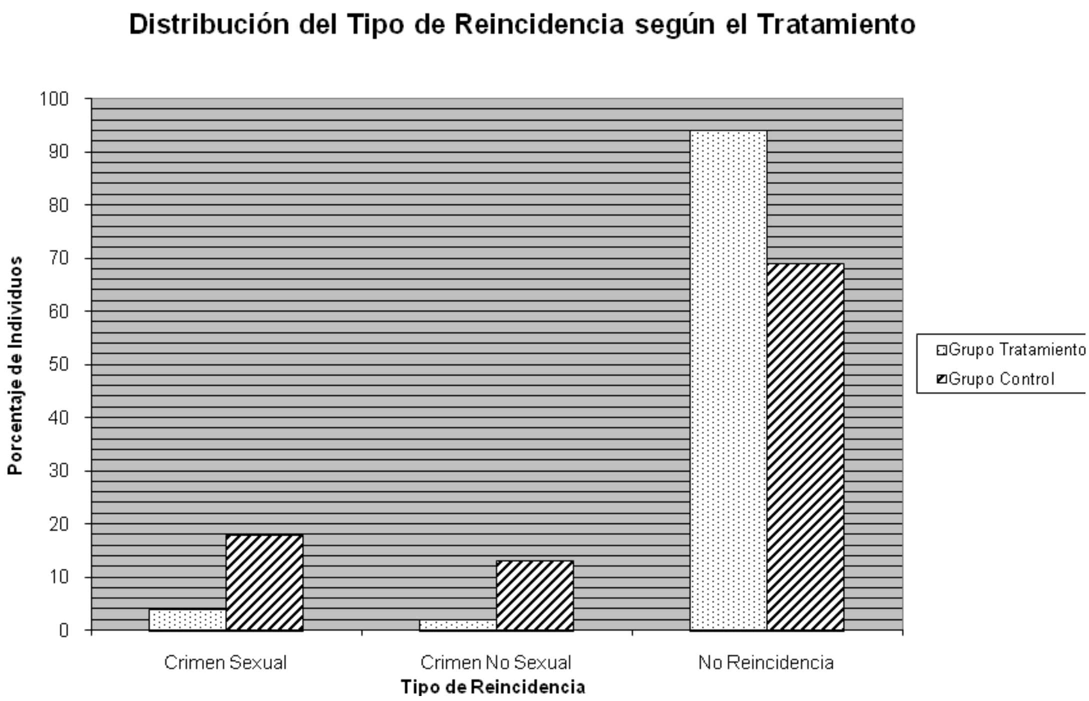
Asignatura	p_i	p_a
Psicoanálisis	**1,2**	**1,2**
Neurofisiología	0,1	**1,3**
Psicología Genética	**- 0,8**	**0,5**
Estadística	0,5	**1**
Total	1	

ERROR 1) Se trata de una variable cualitativa. Por lo tanto, carece de sentido acumular las frecuencias (absolutas, relativas o porcentuales) de los valores de una variable que han sido medidos en escala nominal.

ERROR 2) Las frecuencias relativas son índices que oscilan entre 0 y 1. Por ende, es imposible encontrar valores de p_i como 1,2 y –0,8.

EJERCICIO 5

En Redondo y Garrido (2008) se evalúa la eficacia de cierto tratamiento de delincuentes sexuales que se hallaban en la prisión de Brians, Barcelona. Para ello se seleccionaron dos grupos mediante criterios que permitieron considerarlos equivalentes; el grupo al que se le aplicó el tratamiento consistía de 49 sujetos y el otro, llamado grupo control, de 74. Finalizado el período de tratamiento y una vez cumplida la sentencia, se llevó a cabo un seguimiento de los mismos durante aproximadamente 4 años y se registró si tuvieron reincidencias en delitos sexuales, otros delitos o no tuvieron reincidencias. Los resultados redondeados se exhiben en el siguiente gráfico.



a) Elabore las distribuciones de frecuencias, absolutas, relativas y porcentuales, a partir de la información que presenta el gráfico.
b) ¿Qué tipo de gráfico se ha empleado para representar la información?
c) Mencione la fuente sistemática de variación.
d) Compare ambas distribuciones y extraiga alguna conclusión acerca de los resultados.

El artículo completo se halla en http://www.psicothema.com/pdf/3422.pdf

a) A partir del gráfico podemos leer, en cada grupo, cuáles son los porcentajes de individuos correspondientes a cada tipo de reincidencia.

Así, dentro del grupo que recibió tratamiento, un 4% reincidió en delitos sexuales, un 2% en otros delitos y un 94% no tuvo reincidencias. Dentro del grupo control, un 18% reincidió en delitos sexuales, un 13% en otros delitos y un 69% no tuvo reincidencia.

Estos porcentajes corresponden a las frecuencias porcentuales de cada categoría. Dividiendo cada una por 100, obtenemos la frecuencia relativa en ambos grupos:

Grupo Tratamiento:

Crimen Sexual: $\frac{4}{100} = 0{,}04$

Crimen No Sexual: $\frac{2}{100} = 0{,}02$

No Reincidencia: $\frac{94}{100} = 0{,}94$

Grupo Control:

Crimen Sexual: $\frac{18}{100} = 0{,}18$

Crimen No Sexual: $\frac{13}{100} = 0{,}13$

No Reincidencia: $\frac{69}{100} = 0{,}69$

Para obtener las frecuencias absolutas, en cada grupo, calculamos los porcentajes correspondientes a cada tipo de reincidencia:

<u>Grupo Tratamiento</u>:

Crimen Sexual: 4% de 49, es decir, $0{,}04 \cdot 49 = 1{,}96$
2 sujetos

Crimen No Sexual: 2% de 49, es decir, $0{,}02 \cdot 49 = 0{,}98$
1 sujeto

No Reincidencia: 94% de 49, es decir, $0{,}94 \cdot 49 = 46{,}06$
46 sujetos

<u>Grupo Control</u>:

Crimen Sexual: 18% de 74, es decir, $0{,}18 \cdot 74 = 13{,}32$
13 sujetos

Crimen No Sexual: 13% de 74, es decir, $0{,}13 \cdot 74 = 9{,}62$
10 sujetos

No Reincidencia: 69% de 74, es decir, $0{,}69 \cdot 74 = 51{,}06$
51 sujetos

La siguiente tabla muestra los resultados obtenidos:

	Grupo Tratamiento			Grupo Control		
Tipo de Reincidencia	n_i	p_i	P_i%	n_i	p_i	P_i%
Crimen Sexual	2	0,04	4	13	0,18	18
Crimen No sexual	1	0,02	2	10	0,13	13
No Reincidencia	46	0,94	94	51	0,69	69
	49	1	100	74	1	100

b) Se ha empleado un diagrama de rectángulos adyacentes.
c) La fuente sistemática es el tipo de tratamiento: haberlo recibido o no.
d) El tratamiento parece ser eficaz dado que en esta muestra se observó un porcentaje considerablemente mayor (25% = 94% – 69%) de no reincidencia para el grupo que lo recibió.

EJERCICIO 6

La siguiente tabla de distribución de frecuencias corresponde a un estudio de habilidades numéricas para el cual se ha tomado la cantidad de problemas resueltos por un grupo de niños:

X_i	n_i	n_a	p_i	p_a	$P_{i\%}$	$P_{a\%}$
16 – 20		**150**				**a**
11 – 15				0,7		**b**
6 – 10	30				**c**	
1 – 5					40	
Total		-		-		-

a) Complete la tabla a partir de los datos presentados.
b) Obtenga los puntos medios y los límites exactos de cada uno de los intervalos

a) El tamaño de la muestra es 150 (recordemos que la frecuencia acumulada del valor mayor es *n*); corresponde al 100% de los datos. El valor de **a** es 100.

La frecuencia relativa acumulada correspondiente al intervalo 11 – 15 es 0,7. La frecuencia porcentual acumulada de dicho intervalo es 70:

$$b = 0{,}7 \cdot 100 = 70$$

El valor **c** lo calculamos por medio de una regla de tres simple:

150 niños ———— 100%
30 niños ———— **c** %

$$c = \frac{30 \cdot 100\%}{150} = 20\% \qquad \Rightarrow$$ **c** = 20

Ahora estamos en condiciones de completar las frecuencias relativas y porcentuales acumuladas:

X_i	n_i	n_a	p_i	p_a	$P_{i\%}$	$P_{a\%}$
16 – 20		150		1	**e**	100
11 – 15				0,7	**d**	70
6 – 10	30			0,6	20	60
1 – 5				0,4	40	40
Total		-		-		-

Las frecuencias porcentuales de los dos intervalos superiores se obtienen de la diferencia entre la porcentual acumulada de cada intervalo y la del intervalo inmediato anterior:

d = 70 – 60 = 10

e = 100 – 70 = 30

Conocida la frecuencia porcentual de cada intervalo, es fácil completar el resto de la tabla.

Frecuencias absolutas:

El 40% de 150 es igual a 0,4 · 150 = 60
El 10% de 150 es igual a 0,1 · 150 = 15
El 30% de 150 es igual a 0,3 · 150 = 45

Frecuencias relativas:

Se divide la frecuencia porcentual por 100.

Frecuencias absolutas acumuladas:

Para cada intervalo sumamos su frecuencia absoluta más la absoluta acumulada del valor anterior.

A continuación se muestra la tabla con todos los valores calculados:

X_i	n_i	n_a	p_i	p_a	$P_i\%$	$P_a\%$
16 – 20	45	150	0,3	1	30	100
11 – 15	15	105	0,1	0,7	10	70
6 – 10	30	90	0,2	0,6	20	60
1 – 5	60	60	0,4	0,4	40	40
Total		-		-		-

b) Los límites exactos de cada intervalo representan a los valores máximo y mínimo incluidos en el intervalo y que podrían medirse si se contara con un instrumento de medición perfecta. Para calcularlos basta con restar media unidad de medida (en este caso, 0,5) del límite aparente inferior y sumar media unidad al superior. Así, los límites exactos de los cuatro intervalos serán: 0,5 – 5,5; 5,5 – 10,5; 10,5 – 15,5; 15,5 – 20,5.

El punto medio de cada uno es la semisuma de sus límites exactos:

$$\frac{0{,}5+5{,}5}{2}=3\,;\quad \frac{5{,}5+10{,}5}{2}=8\,;\quad \frac{10{,}5+15{,}5}{2}=13\,;\quad \frac{15{,}5+20{,}5}{2}=18$$

La tabla completa es:

X_i	N_i	n_a	p_i	p_a	$P_i\,\%$	$P_a\,\%$	Puntos Medios	Límites Exactos
16 - 20	45	150	0,3	1	30	100	18	15,5 – 20,5
11 - 15	15	105	0,1	0,7	10	70	13	10,5 – 15,5
6 - 10	30	90	0,2	0,6	20	60	8	5,5 – 10,5
1 – 5	60	60	0,4	0,4	40	40	3	0,5 – 5,5
Total	150	-	1	-	100	-		

EJERCICIO 7

Las siguientes calificaciones han sido obtenidas por 30 adultos en una prueba que evalúa su inteligencia:

100	95	86	97	104	105	110	85	96	102
103	112	98	106	101	104	95	114	89	91
92	115	120	117	99	97	95	100	90	112

a) Presente la información en una distribución de frecuencias (absolutas, relativas, porcentuales con sus correspondientes acumuladas) con intervalos de amplitud 6. Redondee las frecuencias relativas a 4 decimales.
b) Determine los puntos medios y límites exactos de cada uno de los intervalos.
c) Grafique esta distribución con un histograma y un polígono de frecuencias absolutas.
d) Confeccione una nueva tabla que contenga la información en 8 intervalos de amplitud 5, pero presente sólo las frecuencias absolutas.
e) Exhiba en un diagrama de tallo y hojas las calificaciones de los 30 adultos.

a) y b)
Los pasos a seguir para construir una distribución de frecuencias de datos agrupados son:

i) Determinar la amplitud total de los datos. La amplitud total es la diferencia entre el mayor y el menor de los valores observados, es decir:

$$\text{Amplitud Total} = X_{\text{máx}} - X_{\text{mín}}$$

Algunos autores proponen sumarle dos veces la mitad de la unidad de medida (habitualmente esto se reduce a sumarle 1), justificándolo de la siguiente manera: si la variable es continua y se mide en unidades, la asignación del valor, por ejemplo, 85, indicaría que la magnitud es algún valor comprendido entre 84,5 y 85,5. Por lo tanto, la máxima distancia potencial sería la que quedaría entre el mínimo valor cuya observación se informaría con el valor mínimo observado, y el máximo valor cuya observación se informaría con el valor máximo observado.

Así, en nuestra distribución los valores máximo y mínimo observados son 120 y 85, se consideraría, entonces, que la máxima distancia potencial entre los valores sería la comprendida entre 120,5 y 84,5.

$$120{,}5 - 84{,}5 = 36$$

ii) Determinar la amplitud de cada intervalo o el número de intervalos que deseamos contenga nuestra distribución. En este caso la amplitud es conocida, deseamos construir intervalos de amplitud 6. Siendo el total de observaciones 30, podemos formar grupos de 6 observaciones

La amplitud mínima posible de cada intervalo se obtiene dividiendo la amplitud total del conjunto de datos entre el número de intervalos que deseamos:

$$\frac{Amplitud\ Total}{Número\ de\ \mathrm{int}\,ervalos} = Amplitud\ \ de\ cada\ \mathrm{int}\,ervalo$$

De donde, el número de intervalos de la distribución, conocida la amplitud de cada uno, es:

$$Número\ de\ \mathrm{int}\,ervalos = \frac{Amplitud\ Total}{Amplitud\ \ de\ cada\ \mathrm{int}\,ervalo}$$

$$Número\ de\ \mathrm{int}\,ervalos = \frac{36}{6} = 6$$

El número de intervalos que contendrá nuestra distribución de frecuencias es 6.

iii) Determinamos, ahora, los límites de cada intervalo. Si bien no hay normas muy rigurosas al respecto, debemos tener en cuenta las siguientes reglas:

- El intervalo superior debe incluir al mayor valor observado
- El intervalo inferior debe incluir al menor valor observado

Con estas condiciones, haremos, ya que es posible en este caso, que el límite inferior del primer intervalo coincida con el menor valor observado, en este caso, 85. Determinado el límite inferior de este primer intervalo, enumeramos todos los siguientes. Los límites del primero son 85 – 90 (Recordemos que la **amplitud de cada intervalo es la diferencia entre los límites exactos superior e inferior**; en este caso, 90,5 – 84,5 = 6). Al enumerar los demás intervalos, debemos estar seguros de que los intervalos sean continuos y mutuamente excluyentes, (ningún dato debe quedar incluido en más de un intervalo).

iv) A continuación realizamos el conteo de los datos en cada intervalo de clase, obteniendo así las frecuencias absolutas de cada uno de ellos.

Los intervalos con sus respectivas frecuencias se muestran en la siguiente tabla:

X_i	Conteo	n_i
115 – 120	///	3
109 – 114	////	4
103 – 108	~~////~~	5
97 – 102	~~////~~ ///	8
91 – 96	~~////~~ /	6
85 – 90	////	4

Calculamos las frecuencias relativas, porcentuales y las correspondientes acumuladas.

<u>Frecuencias relativas</u>:

Para el intervalo 85 – 90 : $\frac{4}{30} = 0{,}1\widehat{3}$, redondeamos a 0,1333

Para el intervalo 91 – 96 : $\frac{6}{30} = 0{,}2$

Para el intervalo 97 – 102 : $\frac{8}{30} = 0{,}2\widehat{6}$, redondeamos a 0,2667

Para el intervalo 103 – 108 : $\frac{5}{30} = 0{,}1\widehat{6}$, redondeamos a 0,1667

Para el intervalo 109 – 114 : 0,1333

Para el intervalo 115 – 120 : $\frac{3}{30} = 0{,}1$

<u>Frecuencias Porcentuales</u>:

Multiplicamos las frecuencias relativas por 100, obteniendo para cada intervalo: 13,33; 20; 26,67; 16,67; 13,33 y 10.

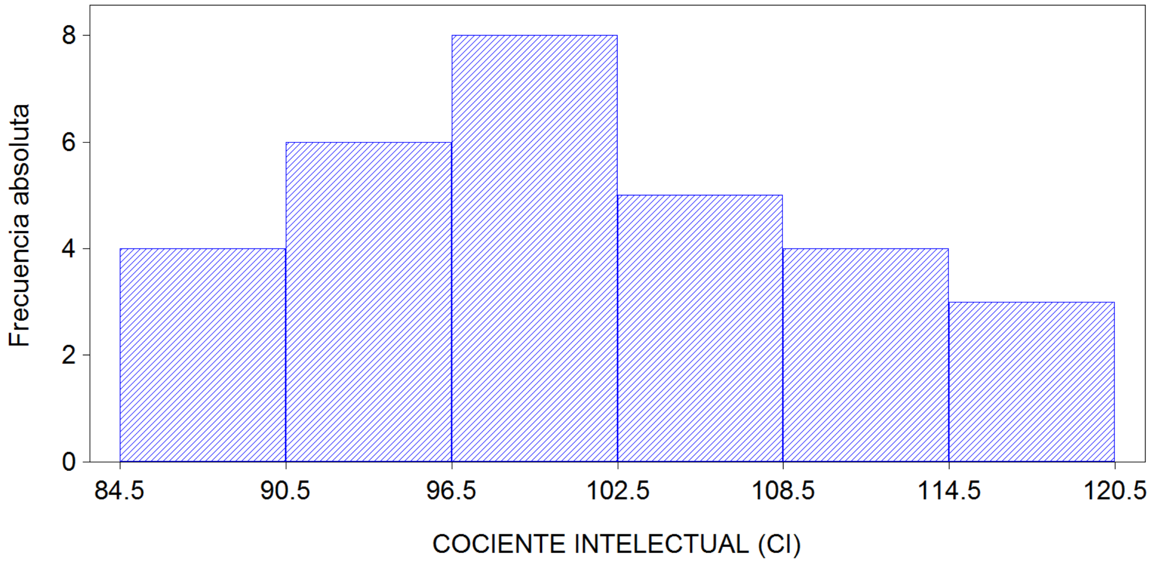
<u>Frecuencias acumuladas</u>:

Las respectivas frecuencias acumuladas se obtienen igual que en los ejercicios anteriores, sumando a la frecuencia de cada intervalo la acumulada del intervalo anterior.

Los resultados se vuelcan en la siguiente tabla:

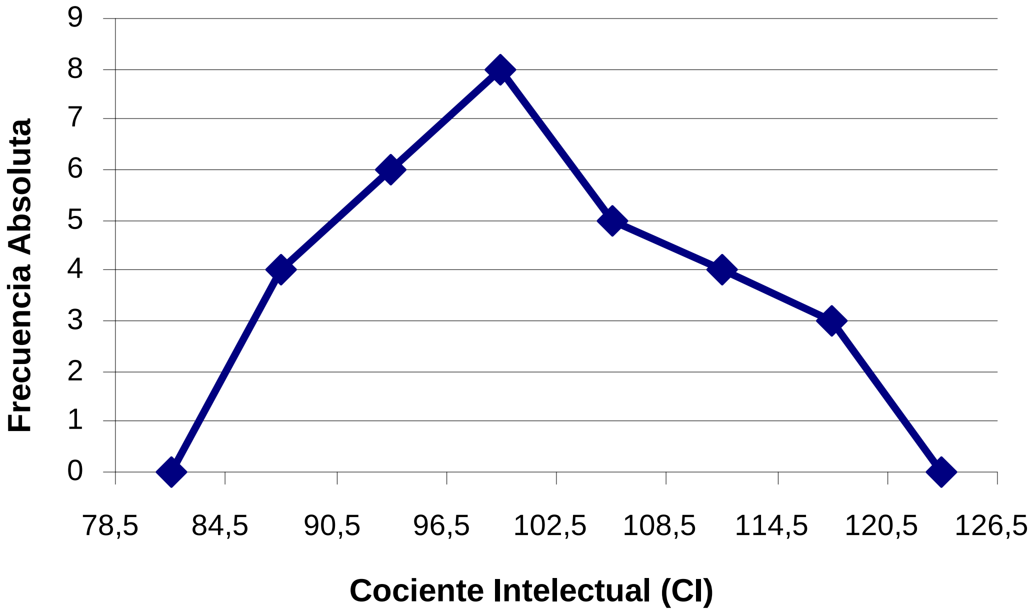
X_i	n_i	n_a	p_i	p_a	P_i%	P_a%	Puntos medios	Límites Exactos
115 – 120	3	30	0,1000	1	10	1	117,5	114,5 – 120,5
109 – 114	4	27	0,1333	0,9000	13,33	90	111,5	108,5 – 114,5
103 – 108	5	23	0,1667	0,7667	16,67	76,67	105,5	102,5 – 108,5
97 – 102	8	18	0,2667	0,6000	26,67	60	99,5	96,5 – 102,5
91 – 96	6	10	0,2000	0,3333	20	33,33	93,5	90,5 – 96,5
85 – 90	4	4	0,1333	0,1333	13,33	13,33	87,5	84,5 – 90,5
	30	-	1	-	100	-		

c) **Histograma**: Para construir un histograma, sobre el eje de las abscisas ubicamos los límites exactos de cada intervalo de clase y sobre el eje de ordenadas las frecuencias. Sobre cada intervalo levantamos un rectángulo cuya altura sea igual a la frecuencia correspondiente. El área y la altura de cada rectángulo, siendo que todos los intervalos tienen la misma amplitud, son proporcionales a la frecuencia (proporción o porcentaje) del respectivo intervalo. Si los rectángulos no tuvieran todos la misma amplitud, el área de cada rectángulo seguirá siendo proporcional a la correspondiente frecuencia, pero no lo serán las alturas.



Polígono de frecuencias: El eje horizontal es idéntico al del histograma. En lugar de barras, marcamos un punto en la gráfica que corresponde al punto medio de cada intervalo y a una altura igual a la frecuencia del mismo. Los puntos así obtenidos se unen mediante líneas rectas. Debemos agregar dos intervalos contiguos al primer y al último de los intervalos de la distribución, con frecuencia cero. La poligonal obtenida anteriormente se extiende cortando al eje horizontal en el punto medio de los intervalos que hemos agregado logrando, así, una figura cerrada o polígono. El área encerrada por dicho polígono es igual al área que encierra al histograma.

También es posible crear el polígono de frecuencias a partir de un histograma. Lo único que hay que hacer es marcar el punto medio de la base superior de cada rectángulo agregando, igual que antes, dos intervalos de frecuencia cero, uno al principio y otro al final de la distribución. Luego unimos los puntos obtenidos.



d) Queremos que nuestra tabla presente la información en 8 intervalos de amplitud 5. En este caso, no podemos construir intervalos de amplitud constante tales que el mayor tenga como límite aparente superior el mayor valor observado, 120, y el intervalo inferior tenga como límite aparente inferior, el menor valor observado, 85, ya que 36 no es múltiplo de 5. En estos casos suele agregarse al conjunto de datos, valores no observados en la muestra, con frecuencia absoluta cero. Para no distorsionar demasiado ninguno de los intervalos extremos estos valores que agregamos se reparten lo más homogéneamente posible entre los dos. Por ejemplo, agregando los valores 83 y 84 en el intervalo inferior, y los valores 121 y 122 en el superior.

X_i	n_i
118 – 122	1
113 – 117	3
108 – 112	3
103 – 107	4
98 – 102	7
93 – 97	6
88 – 92	4
83 – 87	2

e) **Diagrama de tallo y hoja**: para construir un diagrama de tallo y hoja, debemos separar cada puntuación en dos partes: el primer o primeros dígitos, que llamaremos *tallo*, y el dígito o dígitos restantes, que llamaremos *hoja*. Para ello:

i) Identificamos los valores máximo y mínimo observados.

ii) Decidimos el número de tallos distintos más apropiado para nuestra distribución.

iii) Se escriben todos los tallos en una columna, ordenados en forma creciente, de arriba hacia abajo.

iv) Se escribe cada hoja junto al tallo que le corresponda, ordenadas según su valor.

8	569
9	01255567789
10	001234456
11	022457
12	0

EJERCICIO 8

Un grupo de 20 niños de una escuela pública de Santiago del Estero han sido evaluados con una prueba de ortografía. Estos son los resultados:

2,5	5	6,5	9	4,5
5,5	8	6	5,5	10
3,5	8,5	7	5	4
4,5	9,5	6	7,5	6,5

a) Obtenga la distribución de frecuencias completa con 4 intervalos.
b) Determine los puntos medios y límites exactos de cada uno de los intervalos.
c) Grafique esta distribución con un polígono de frecuencias absolutas.

d) Dibuje un polígono de frecuencias porcentuales acumuladas.

a) y b)
La máxima distancia entre los valores máximo y mínimo observados es:

10,05 – 2,45 = 7,6

Para obtener la amplitud de cada intervalo, dividimos la amplitud total por el número de intervalos, 4:

$$\frac{7{,}6}{4} = 1{,}9$$

La amplitud de cada intervalo es 1,9.

Así, los límites de cada intervalo son: 2,5 – 4,3; 4,4 – 6,2; 6,3 – 8,1; 8,2 – 10

A veces resulta más sencillo comenzar por ubicar los límites exactos en vez de trabajar con los límites aparentes. Veamos como sería con el intervalo inferior:

Amplitud de cada intervalo = Límite Exacto Superior (LES) – Límite Exacto Inferior (LEI)

1,9 = LES – 2,45

de donde, 1,9 + 2,45 = LES

4,35 = LES

Los límites exactos son: 2,45 – 4,35; los aparentes, entonces, son: 2,5 – 4,3

Determinados los cuatro intervalos, realizamos el conteo de los datos:

X_i	Conteo	n_i
8,2 -10	////	4
6,3 - 8,1	$\cancel{////}$	5
4,4 - 6,2	$\cancel{////}$ ///	8
2,5 - 4,3	///	3

Frecuencias absolutas acumuladas:

Sumamos para cada valor su frecuencia absoluta más la absoluta acumulada del valor anterior, obteniendo los valores 3, 11, 16 y 20.

Frecuencias relativas:

Dividimos cada frecuencia absoluta por el tamaño de la muestra, es decir, 20:

$$\frac{4}{20} = 0{,}2\,; \quad \frac{5}{20} = 0{,}25; \quad \frac{8}{20} = 0{,}4\,; \quad \frac{3}{20} = 0{,}15$$

La suma de las cuatro frecuencias relativas es 1.

Frecuencias relativas acumuladas:

Para cada intervalo sumamos su frecuencia absoluta más la absoluta acumulada del valor anterior: 0,15 ; 0,55 ; 0,80 ; 1

Frecuencias porcentuales:

Multiplicamos las frecuencias relativa y relativa acumulada por 100.

Los resultados se muestran en la siguiente tabla:

X_i	n_i	n_a	p_i	p_a	P_i%	P_a%
8,2 – 10	4	20	0,20	1,00	20	100
6,3 – 8,1	5	16	0,25	0,80	25	80
4,4 – 6,2	8	11	0,40	0,55	40	55
2,5 – 4,3	3	3	0,15	0,15	15	15

Los límites exactos los obtenemos restando del límite aparente inferior (LAI) y sumando al límite aparente superior (LAS), media unidad de medida, es decir, 0,05:

$$LEI = LAI - 0{,}05$$

$$LES = LAS + 0{,}05$$

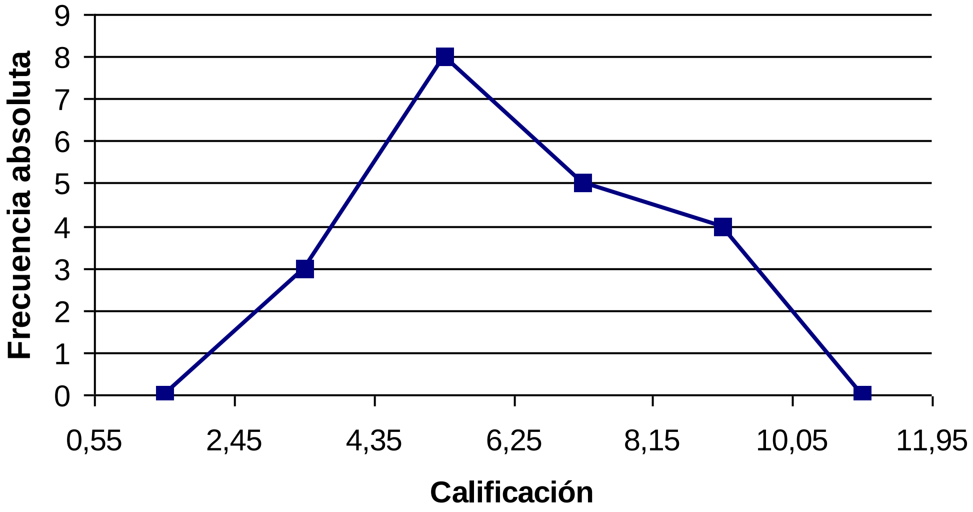
Los puntos medios son:

$$\frac{2{,}45+4{,}35}{2}=3{,}4\ ;\quad \frac{4{,}35+6{,}25}{2}=5{,}3\ ;\quad \frac{6{,}25+8{,}15}{2}=7{,}2\ ;\quad \frac{8{,}15+10{,}05}{2}=9{,}1$$

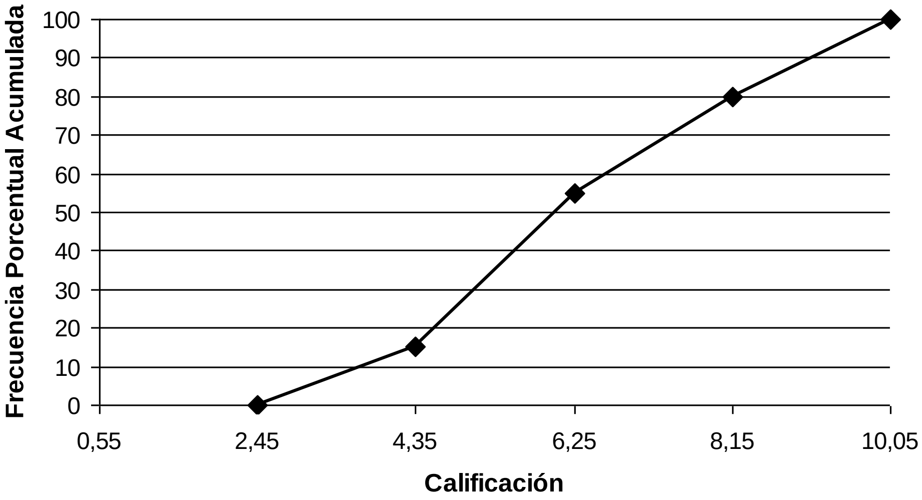
La tabla completa es:

X_i	n_i	n_a	p_i	p_a	P_i%	P_a%	Puntos medios	Límites Exactos
8,2 – 10	4	20	0,20	1,00	20	100	9,1	8,15 – 10,05
6,3 – 8,1	5	16	0,25	0,80	25	80	7,2	6,25 – 8,15
4,4 – 6,2	8	11	0,40	0,55	40	55	5,3	4,35 – 6,25
2,5 – 4,3	3	3	0,15	0,15	15	15	3,4	2,45 – 4,35
	20	-	1	-	100	-		

c) **Polígono de frecuencias absolutas**: marcamos en el gráfico los puntos que corresponden al punto medio de cada intervalo y altura igual a la frecuencia del mismo. Agregamos dos intervalos contiguos al primer y al último de los intervalos de la distribución, con frecuencia cero: 0,55 – 2,45 y 10,05 – 11,95. Unimos todos los puntos mediante líneas rectas.



d) **Polígono de frecuencias porcentuales acumuladas**: El eje de abscisas se construye igual que en el gráfico anterior, pero sobre el eje de ordenadas ubicamos las frecuencias acumuladas (en este tipo de gráfico se suelen utilizar las frecuencias acumuladas porcentuales). Sobre el límite exacto superior de cada intervalo ubicamos un punto cuya altura sea la frecuencia acumulada de dicho intervalo. Unimos mediante líneas rectas los puntos obtenidos.



EJERCICIO 9

Qué gráfico utilizaría para representar la distribución de frecuencias de:

a) La cantidad de materias aprobadas por los alumnos de una comisión.
b) La orientación política de los habitantes de la provincia de Río Negro.
c) Los resultados obtenidos por un alumno en una batería de tests.
d) La opinión acerca del grado de satisfacción con el empleo actual de los argentinos.
e) Tiempo de reacción frente a un estímulo visual de un grupo de 60 niños en etapa escolar.
f) La opinión que tienen los graduados y los estudiantes sobre el plan de estudios vigente. Comparar ambas opiniones.

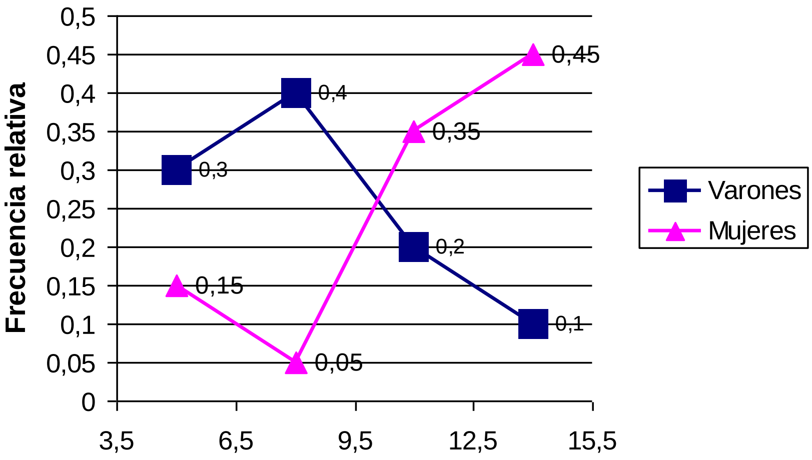
a) Diagrama de barras.
b) Pictograma / Diagrama de rectángulos.
c) Perfil ortogonal.
d) Diagrama de rectángulos.
e) Histograma / Polígono de frecuencias.
f) Diagrama de rectángulos adyacentes.

EJERCICIO 10

Con el fin de estudiar la relación existente entre el género y el índice de religiosidad un investigador ha seleccionado aleatoriamente a 80 mujeres y 20 varones. Aquí reproducimos los resultados obtenidos por ambos grupos. Dibuje un polígono de frecuencias comparando ambas distribuciones y comente los resultados.

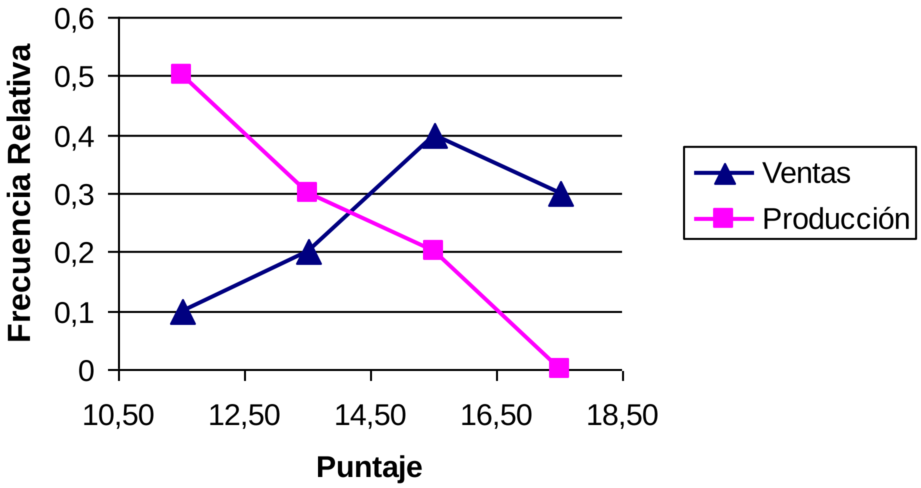
Puntaje	Varones	Mujeres
13 - 15	2	36
10 - 12	4	28
7 - 9	8	4
4 - 6	6	12

Se trata de muestras de tamaños diferentes, es por este motivo que el gráfico no puede ser realizado con frecuencias absolutas. Estos datos muestran una tendencia de las mujeres a obtener índices de religiosidad más elevados que los varones.



EJERCICIO 11

El siguiente diagrama corresponde a los resultados de una prueba administrada a 16 empleados de ventas y 52 empleados del área de producción de una empresa que han respondido a la escala de Extroversión de un test de personalidad; mayores puntajes significan mayor extroversión.



a) Diseñe una tabla de frecuencias absolutas y porcentuales con sus correspondientes frecuencias acumuladas. Incluya los límites exactos y aparentes.
b) i) Cómo se denomina el gráfico empleado.
ii) Por qué se aplicó dicho gráfico.
iii) Por qué se ha utilizado la frecuencia relativa para su confección.
c) ¿Qué conclusión podría extraer de los resultados obtenidos?

a) Del diagrama leemos los límites exactos de cada intervalo, los puntos medios se obtienen hallando la semisuma de los mismos:

Primer intervalo: 10,50 – 12,50. Punto medio: $\frac{10{,}50+12{,}50}{2}=11{,}5$

Segundo intervalo: 12,50 – 14,50. Punto medio: $\frac{12{,}50+14{,}50}{2}=13{,}5$

Tercer intervalo: 14,50 – 16,50. Punto medio: $\frac{14{,}50+16{,}50}{2}=15{,}5$

Cuarto intervalo: 16,50 – 18,50. Punto medio: $\frac{16{,}50+18{,}50}{2}=17{,}5$

Recordemos que los límites exactos se obtienen restando media unidad de medida (en este caso, 0,5) del límite aparente inferior y sumando media unidad al superior. Realizando el proceso inverso obtenemos los ***límites aparentes***:

Para el primer intervalo: 10,5 + 0,5 = 11

12,5 – 0, 5 = 12

Se procede igual para el resto de los intervalos: 11 – 12 ; 13 – 14 ; 15 – 16 ; 17 – 18

<u>Frecuencias relativas</u>:

Están representadas por la altura de cada punto medio. Las mismas, para cada intervalo, son:

Ventas: 0,1; 0,2; 0,4; 0,3

Producción: 0,5; 0,3; 0,2; 0,1

A partir de las frecuencias relativas, multiplicando por 100 cada una de ellas, obtenemos las porcentuales y, luego, calculando los respectivos porcentajes sobre el total de cada grupo, las absolutas.

<u>Frecuencias porcentuales</u>:

Ventas: 10; 20; 40; 30

Producción: 50; 30; 20; 0

<u>Frecuencias absolutas</u>:

Ventas: 10% de 16 = 1,6 $\Longrightarrow$ **2**

20% de 16 = 3,2 $\Longrightarrow$ **3**

40% de 16 = 6,4 $\Longrightarrow$ **6**

30% de 16 = 4,8 $\Longrightarrow$ **5**

Producción: 50% de 52 = **26**

30% de 52 = 15,6 $\Longrightarrow$ **16**

20% de 52 = 10,4 $\Longrightarrow$ **10**

Las respectivas ***acumuladas*** se obtienen sumando, para cada intervalo, la frecuencia correspondiente a la acumulada del valor anterior.

Los resultados se muestran en la siguiente tabla:

	Ventas				Producción					
X_i	n_i	n_a	P_i%	P_a%	n_i	n_a	P_i%	P_a%	Puntos medios	Límites Exactos
11 – 12	2	2	10	10	26	26	50	50	11,5	10,5 – 12,5
13 – 14	3	5	20	30	16	42	30	80	13,5	12,5 – 14,5
15 – 16	6	11	40	70	10	52	20	100	15,5	14,5 – 16,5
17 – 18	5	16	30	100	0	52	0	100	17,5	16,5 – 18,5
	16		100		52		100			

b) Se ha empleado un polígono de frecuencias ya que se pretendía comparar distribuciones referidas a una variable cuantitativa continua. Se ha utilizado la frecuencia relativa porque las muestras son de diferente tamaño.

c) En el área de Ventas son más frecuentes los puntajes altos de extroversión que en el de Producción; por lo que se concluye que para la muestra considerada los empleados de la sección Ventas son en general más extrovertidos que los del área de Producción.

REFERENCIAS

Piaget, J. y Inhelder, B. (1970). *El Desarrollo de las Cantidades en el Niño.* Introducción. Editorial Nova Terra. Barcelona.

Redondo, S. y Garrido, V. (2008). Efficacy of a psychological treatment for sex offenders. *Psicothema, 20*, 1, 4 - 9.

Simón, T., Ruiz Gallego-Largo, T. y Suengas, A. (2009). Memoria y envejecimiento: recuerdo, reconocimiento y sesgo positivo. *Psicothema, 21*, 3, 409 - 415.

PRÁCTICA III

EJERCITACIÓN

MEDIDAS DE POSICIÓN

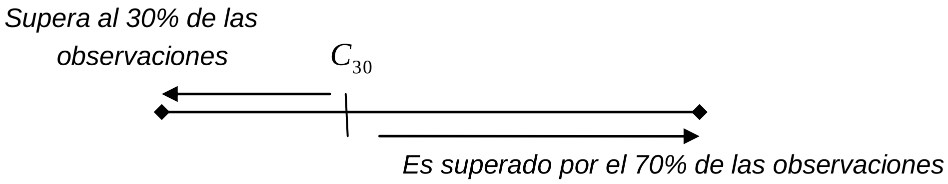
Recordemos que las medidas de posición son índices que brindan información acerca de la situación de una puntuación con respecto a un grupo, utilizando a éste como marco de referencia. Estas medidas son los **cuantiles**.

Los cuantiles son valores que se definen en función del *porcentaje de observaciones a las que superan*.

Distintos cuantiles son: los centiles, los deciles, los cuartiles y los quintiles.

Recordemos la definición de cada uno de ellos para **variables cuantitativas continuas**.

Los **centiles** son 99 valores que determinan 100 intervalos en el recorrido de la variable, cada uno de los cuáles contiene el 1% de las observaciones. En símbolos, utilizamos la notación **C*k***, donde ***k*** recibe el nombre de **rango percentilar**. Así, C_{30}, por ejemplo, es la puntuación que supera al 30% de las observaciones y es superada por el 70% de las mismas. Podemos representar gráficamente esta situación de la siguiente manera:



El valor ***k*** puede generalizarse a valores no enteros.

Análogamente definimos los deciles y los cuartiles.

Los **deciles** son 9 valores que determinan 10 intervalos en el recorrido de la variable, cada uno de los cuáles contiene el 10% de las observaciones. Utilizamos la notación D_k.

Los **cuartiles** son 3 valores que determinan 4 intervalos en el recorrido de la variable, cada uno de los cuáles contiene el 25% de las observaciones. Ellos son Q_1, Q_2 y Q_3.

Existe una equivalencia entre los distintos cuantiles. Por ejemplo, el C_{30} es equivalente al D_3, ya que ambos representan la puntuación que supera al 30% de las observaciones y es superada por el 70% de las mismas.

También, son equivalentes las puntuaciones C_{50}, D_5 y Q_2. Gracias a esta equivalencia, las fórmulas de los cuantiles se resumen en la de los centiles correspondientes al cuantil que se quiere calcular.

La fórmula que utilizamos para su cálculo es:

$$C_k = Li + \frac{I}{n_i} \cdot \left(\frac{k \cdot n}{100} - n_a \right)$$

donde:

L_i es el límite exacto inferior del intervalo crítico
I es la amplitud de los intervalos
n_i es la frecuencia absoluta del intervalo crítico
k es el porcentaje de observaciones inferiores a C_k
n es el número de observaciones hechas
n_a es la frecuencia acumulada hasta L_i

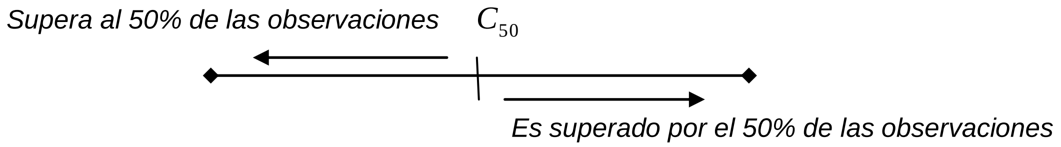
EJERCICIO 1

Un investigador pretende analizar la posible relación entre soledad y uso de Internet. Para tal fin interrogó a 35 personas que viven solas y a 25 que viven con sus familias acerca de la cantidad de minutos que suelen estar conectadas en una sesión típica durante el fin de semana. Ambas muestras fueron homogeneizadas en la edad y el nivel socioeconómico. Los datos son los siguientes:

Tiempo (en minutos)	Personas que viven solas	Personas que viven con su familia
60 - 74	3	1
45 - 59	11	3
30 - 44	10	8
15 - 29	6	7
0 - 14	5	6

Calcule el D_5 en cada grupo. ¿Qué indica este resultado respecto de la relación entre la soledad y el uso de Internet?

Siendo el D_5 equivalente al C_{50}, calculamos el C_{50} en cada grupo:



a) Comenzamos por calcular el porcentaje de observaciones que supera dicha puntuación en el grupo de las personas que viven solas. El 50% de 35 es:

$$\frac{k \cdot n}{100} = \frac{50 \cdot 35}{100} = 17{,}5$$

b) Determinación del **intervalo crítico**: El intervalo crítico es el intervalo que contiene a la puntuación que buscamos, en este caso, al D_5 o C_{50}. Para ello, necesitamos calcular las frecuencias absolutas acumuladas y buscar en esta columna el intervalo que contiene acumuladas 17,5 observaciones; 17,5 se acumula en el interior del intervalo 30 – 40; sus límites exactos son 29,5 y 44,5.

Tiempo (en minutos)	Personas que viven solas	Frecuencia absoluta acumulada
60 - 74	3	33
45 - 59	11	32
30 - 44	**10**	**21**
15 - 29	6	11
0 - 14	5	5

Utilizamos la fórmula correspondiente al cálculo de centiles para variables cuantitativas continuas.

$$C_{50} = 29{,}5 + \frac{15}{10} \cdot \left(\underbrace{\frac{50 \cdot 35}{100}}_{17{,}5} - 11 \right) \quad \Rightarrow \quad C_{50} = D_5 = 39{,}25$$

Personas que viven solas: D_5 = 39,25 min.

Análogamente calculamos el D_5 para las personas que viven con su familia.

Tiempo (en minutos)	Personas que viven con su familia	Frecuencia absoluta acumulada
60 – 74	1	25
45 – 59	3	24
30 – 44	8	21
15 – 29	**7**	**13**
0 – 14	6	6

$$\frac{k \cdot n}{100} = \frac{50 \cdot 25}{100} = 12{,}5$$

Buscamos en la tabla, en la columna de frecuencias absolutas acumuladas, el intervalo que contiene a 12,5 observaciones. Dicho intervalo es 15 – 29. Calculamos el C_{50}:

$$C_{50} = 14{,}5 + \frac{15}{7} \cdot (12{,}5 - 6) \quad \Rightarrow \quad C_{50} = D_5 = 28{,}43$$

Personas que viven con su familia: D_5 = 28,43 min.

La diferencia entre los valores del quinto decil indica una tendencia de las personas que viven solas a estar más conectadas a Internet que aquellos que viven con sus familias. Esta afirmación en principio es válida sólo para los individuos de la muestra. Sin embargo, es claro que el objetivo de este tipo de estudios no es concluir sobre los individuos en particular que participaron de la investigación sino en general sobre una población más amplia. Habría que ver si la diferencia entre los deciles se explica, al menos parcialmente, por vivir o no con una familia o es casual (por azar); es decir: cualquier otra muestra de individuos ¿arrojaría resultados semejantes? Esto se vincula directamente con el tema de la variabilidad: si la diferencia entre los deciles variara poco de muestra en muestra, con la mayoría de las muestras se afirmaría lo mismo: que las personas que viven solas tienden a pasar más tiempo conectadas a Internet; con lo cual, vivir solo o con la familia sería una fuente sistemática de variación; pero si los deciles variaran mucho, muestras distintas podrían arrojar resultados contradictorios y la afirmación perdería validez general.

EJERCICIO 2

Replicando un clásico estudio empírico de la Psicología Social llevado a cabo por Rosenhan (1988), un investigador decidió infiltrar a "falsos pacientes" en distintas clínicas psiquiátricas de EEUU. Su objetivo fue evaluar la reacción del personal al contacto verbal iniciado por el paciente. Los pseudo-pacientes abordaron a psiquiatras y enfermeros de las respectivas instituciones en un encuentro que parecía casual y les formularon un máximo de seis preguntas secuenciadas referidas a su "diagnóstico" y su "pronóstico". El investigador consideró como variable estadística a la cantidad de preguntas respondidas por los profesionales en ese encuentro. Estos son los resultados (datos ficticios):

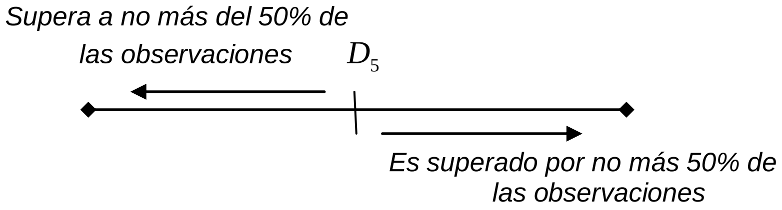
Cantidad de preguntas respondidas	Enfermeros	Psiquiatras
0	1	2
1	3	4
2	6	2
3	8	1
4	9	0
5	7	1
6	1	0

La hipótesis del investigador sostiene que el contacto de los psiquiatras con los pacientes es más frágil que el de los enfermeros. Calcule el segundo cuartil de ambas distribuciones y extraiga conclusiones al respecto.

El estudio se realizó sobre 35 enfermeros y 10 psiquiatras. En este caso, la variable estadística es una variable cuantitativa discreta. Debemos utilizar la **definición de centil extendida a variables cuasicuantitativas o discretas**.

Definición: "Los centiles son 99 valores de la variable que determinan 100 intervalos (a los que no pertenece) en el recorrido de la variable de modo que como máximo la centésima parte de las observaciones pertenecen a cada intervalo".

El segundo cuartil, Q_2, de acuerdo a la definición, es la puntuación que *supera a no más del* 50% de las observaciones y que es *superada por no más del* 50% de las mismas.



Para encontrar dicha puntuación debemos calcular las *frecuencias porcentuales acumuladas* en ambos sentidos.

Cantidad de preguntas respondidas	Enfermeros	f_a%	g_a%
0	1	2,86	100
1	3	11,43	97,14
2	6	**17,14**	88,57
3	8	51,43	71,43
4	9	77,14	**48,57**
5	7	97,14	22,86
6	1	100	2,86

El valor que cumple la condición de superar a no más del 50% de las observaciones y ser superado por no más del 50% de las mismas es 3 (el 17,14% de los enfermeros respondió menos de 3 preguntas y el 48,57% respondió más de 3 preguntas).

Grupo Enfermeros: Q_2 = 3 preguntas respondidas.

Realizamos el mismo procedimiento para el grupo de psiquiatras.

Cantidad de preguntas respondidas	Psiquiatras	f_a%	g_a%
0	2	**20**	100
1	4	60	80
2	2	80	**40**
3	1	90	20
4	0	90	10
5	1	100	10
6	0	100	0

El valor que cumple la misma condición en el grupo de los psiquiatras es 1 (el 20% de los psiquiatras respondió menos de 1 pregunta y el 40% más de 1 pregunta).

Grupo Psiquiatras: Q_2 = 1 pregunta respondida.

Los resultados no difieren mucho de los que se hubieran obtenido de haber tratado a la variable como continua según lo explican Botella, León y San Martín (1993, pág 87) – 3,44 y 1,25 - pero son más exactos ya que se usa toda la información original. Cuando los datos se agrupan en intervalos se pierde información. Esto era útil para facilitar los cálculos cuando no se disponía de medios computacionales como los de hoy en día, donde ya no se justifica.

La diferencia encontrada al comparar los segundos cuartiles está en la misma dirección de la hipótesis formulada por el investigador. Sin embargo, se debe ser prudente en tanto que esta afirmación tiene un alcance limitado en su posibilidad de generalización. Al igual que en la respuesta del Ejercicio 1, sólo describen los datos pertenecientes a la muestra tomada por el investigador.

EJERCICIO 3

Un profesor considera que deben promocionar la asignatura los alumnos que obtuvieron un promedio de notas en ambos parciales superior al centil 90. El profesor utiliza la puntuación 35 para determinar la categoría de “aprobado”. La siguiente es la distribución de frecuencias de los promedios de las notas de dicho profesor:

X_i	n_i
81 - 100	6
61 - 80	25
41 - 60	38
21 - 40	16
1 - 20	5

En función de los datos de la tabla:

a) Calcule el porcentaje de alumnos que desaprobaron la materia.
b) Calcule cuál será la puntuación mínima para promocionar.

Se trata nuevamente de una variable cuantitativa continua (Puntaje promedio obtenido por cada alumno en ambos parciales).

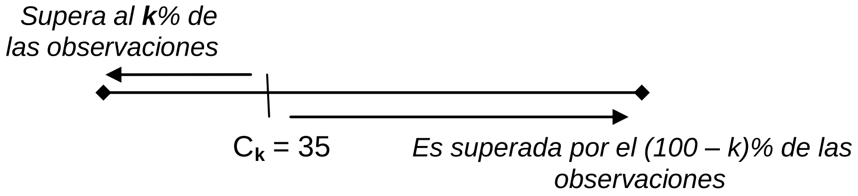
a) Queremos calcular el porcentaje de alumnos que desaprobaron la materia. El profesor utiliza la puntuación 35 para determinar la categoría de “aprobado”, debemos buscar, entonces, el porcentaje de observaciones que deja debajo de sí la puntuación 35.

Los pasos son los siguientes:

- Determinar el intervalo crítico: en este caso, conocemos la puntuación, por lo tanto, el intervalo crítico será aquél que incluye la puntuación 35. Dicho intervalo es 21 – 40.

X_i	n_i	Frecuencia absoluta acumulada
81 - 100	6	90
61 - 80	25	84
41 - 60	38	59
21 - 40	**16**	**21**
1 - 20	5	5

- Utilizamos nuevamente la fórmula de los centiles pero, ahora, la incógnita es ***k***:



$$35 = 20{,}5 + \frac{20}{16}\cdot\left(\frac{k\cdot 90}{100} - 5\right) \quad\Rightarrow\quad k = \left[\ (35 - 20{,}5) : \frac{20}{16} + 5\ \right]\cdot\frac{100}{90} = 18{,}\overset{\frown}{4}$$

El porcentaje de alumnos que desaprobaron la materia 18,44%

b) El centil 90 es la puntuación que supera al 90% de las observaciones y es superada por el 10% de las mismas.

El 90% de las observaciones es

$$\frac{k\cdot n}{100} = \frac{90\cdot 90}{100} = 81$$

El intervalo crítico es el intervalo 61 – 80.

X_i	n_i	Frecuencia absoluta acumulada
81 - 100	6	90
61 - 80	**25**	**84**
41 - 60	38	59
21 - 40	16	21
1 - 20	5	5

Calculamos el centil 90:

$$C_{90} = 60{,}5 + \frac{20}{25} \cdot (81 - 59) \quad \Rightarrow \quad C_{90} = 78{,}1$$

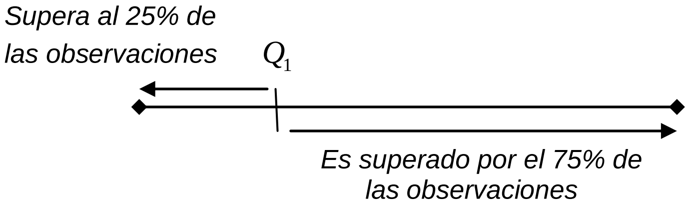
C_{90} = 78,1

EJERCICIO 4

Una institución que asiste a drogadependientes se interesa por conocer el nivel de ansiedad motora de sus pacientes. Los profesionales deciden administrar la prueba ISRA de Miguel Tobal y Cano Vindel (1998). La misma utiliza el primer cuartil para calificar “ansiedad baja”; el tercer cuartil para calificar “ansiedad alta” y el centil 99 para la “ansiedad extrema”. Calcule las puntuaciones que delimitan estas categorías.

Puntaje	n_i
106 – 119	10
92 – 105	45
78 – 91	36
64 – 77	18
50 – 63	20
36 – 49	15

Cálculo del primer cuartil:



El Q_1 es equivalente al C_{25}.

El 25% de las observaciones es $\frac{k \cdot n}{100} = \frac{25 \cdot 144}{100} = 36$

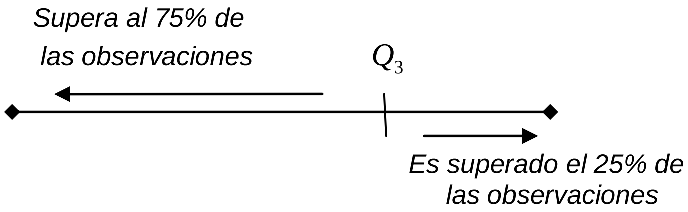
El intervalo crítico es el intervalo 64 – 77.

Puntaje	n_i	Frecuencia absoluta acumulada
106 - 119	10	144
92 - 105	45	134
78 - 91	36	89
64 - 77	**18**	**53**
50 - 63	20	35
36 - 49	15	15

Calculamos el centil 25:

$$C_{25} = 63{,}5 + \frac{14}{18} \cdot (36 - 35) \quad \Rightarrow \quad Q_1 = 64{,}2\widehat{7}$$

Cálculo del tercer cuartil:



El Q_3 es equivalente al C_{75}.

El 75% de las observaciones es $\frac{k \cdot n}{100} = \frac{75 \cdot 144}{100} = 108$

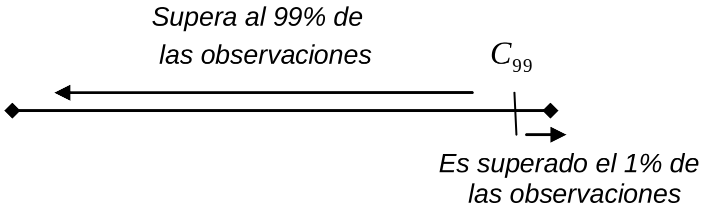
El intervalo crítico es el intervalo 92 – 105.

Puntaje	n_i	Frecuencia absoluta acumulada
106 - 119	10	144
92 - 105	**45**	**134**
78 - 91	36	89
64 - 77	18	53
50 - 63	20	35
36 - 49	15	15

Calculamos el centil 75:

$$C_{75} = 91{,}5 + \frac{14}{45} \cdot (108 - 89) \quad \Rightarrow \quad Q_3 = 97{,}4\widehat{1}$$

Cálculo del centil 99:



El 99% de las observaciones es $\frac{k \cdot n}{100} = \frac{99 \cdot 144}{100} = 142{,}56$

El intervalo crítico es el intervalo 106 – 119.

Puntaje	n_i	Frecuencia absoluta acumulada
106 - 119	**10**	**144**
92 - 105	45	134
78 - 91	36	89
64 - 77	18	53
50 - 63	20	35
36 - 49	15	15

Calculamos el centil 99:

$$C_{99} = 105{,}5 + \frac{14}{10} \cdot (142{,}56 - 134) \quad \Rightarrow \quad C_{99} = 117{,}484$$

Ansiedad Leve **Q_1 = 64,28**	**Ansiedad Moderada** **Q_3 = 97,41**	**Ansiedad Extrema** **C_{99} = 117,48**

EJERCICIO 5

Según un test que evalúa la variable Depresión, los pacientes que obtienen una puntuación superior a 40 presentan riesgo suicida. La siguiente es la distribución de frecuencias de los puntajes obtenidos en el test por 45 mujeres y 60 varones diagnosticados como depresivos, que concurren a un Centro de Salud del Gran Buenos Aires.

Puntaje	Mujeres	Varones
43,5 – 50,5	4	5
36,5 – 43,5	10	2
29,5 – 36,5	14	11
22,5 – 29,5	7	24
15,5 – 22,5	10	18

¿Cuál de los dos grupos presenta el mayor porcentaje de pacientes con riesgo suicida?

Queremos calcular el porcentaje de observaciones que supera la puntuación 40. El intervalo crítico será aquél que incluye dicha puntuación.

<u>Grupo de Mujeres</u>:

El intervalo crítico es 36,5 – 43,5.

Puntaje	Mujeres	Frecuencia absoluta acumulada
43,5 – 50,5	4	45
36,5 – 43,5	**10**	**41**
29,5 – 36,5	14	31
22,5 – 29,5	7	17
15,5 – 22,5	10	10

Utilizamos nuevamente la fórmula de los centiles donde la incógnita es ***k***:

$$40 = 36{,}5 + \frac{7{,}1}{10} \cdot \left(\frac{k \cdot 45}{100} - 31 \right) \quad \Rightarrow \quad k = \left[(40 - 36{,}5) : \frac{7{,}1}{10} + 31 \right] \cdot \frac{100}{45} \approx 80$$

La puntuación 40 deja por debajo de sí el 80% de las observaciones, por lo tanto, el 20% de las mismas superan a dicha puntuación, porcentaje que representa a los pacientes que presentan riesgo suicida.

<u>Grupo de Varones</u>:

El intervalo crítico es 36,5 – 43,5.

Puntaje	Varones	Frecuencia absoluta acumulada
43,5 – 50,5	5	60
36,5 – 43,5	**2**	**55**
29,5 – 36,5	11	53
22,5 – 29,5	24	42
15,5 – 22,5	18	18

$$40 = 36{,}45 + \frac{7{,}1}{2} \cdot \left(\frac{k \cdot 60}{100} - 53 \right) \quad \Rightarrow \quad k = \left[\left(40 - 36{,}45\right) : \frac{7{,}1}{2} + 53 \right] \cdot \frac{100}{60} = 90$$

La puntuación 40 deja por debajo de sí el 90% de las observaciones, y es superada por el 10% de las mismas (porcentaje de pacientes con riesgo suicida).

En esta muestra, las mujeres tienen un mayor porcentaje de riesgo suicida.
Mujeres: 20% - Varones: 10%

EJERCICIO 6

El gerente de una empresa se pregunta si los años de experiencia de un obrero en su puesto favorecen el cumplimiento de las normas de seguridad y mantenimiento de las maquinarias. Para ello decide:

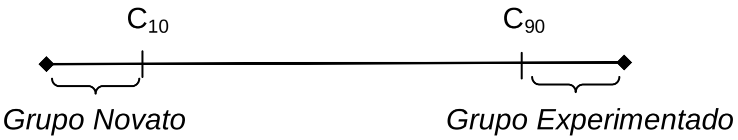
1) *Seleccionar dos grupos bien diferenciados*. El grupo de los novatos que estará compuesto por aquellos operarios cuyos años de experiencia no superen el centil 10 y el grupo experimentado que estará compuesto por los que superen el centil 90. Sabiendo que la empresa emplea a 200 operarios ¿cuántas personas integrarán cada uno de los grupos?

2) *Evaluar a los grupos y analizar los resultados*. Ambos grupos respondieron una prueba destinada a evaluar el conocimiento y cumplimiento de las normas de seguridad y mantenimiento de las maquinarias. Los resultados obtenidos son los siguientes:

Puntaje	Grupo Novato	Grupo Experimentado
17 – 20	2	1
13 – 16		3
9 – 12	8	
5 – 8	5	6
1 – 4	2	2
Total		

a) Complete la tabla a partir de los datos obtenidos en 1).
b) Obtenga los cuartiles del grupo novato.
c) Si los cuartiles para el grupo experimentado son los siguientes: Q_1= 6,5; Q_2=9,5; Q_3=12, ¿cómo interpretaría estos resultados comparando ambos grupos para responder a la pregunta del gerente?

1)



El grupo de los novatos está compuesto por el 10% de las observaciones que no supera el C_{10} y el grupo de los experimentados por el 10% de las mismas que superan al C_{90}. Siendo 200 los operarios que pertenecen a la empresa, cada grupo estará formado por 20 operarios (10% de 200).

2)

a) El total de cada grupo es 20 operarios. Dentro del grupo novato, la cantidad de operarios que obtuvo un puntaje perteneciente al intervalo 13 – 16 es:

$$20 - (2 + 8 + 5 + 2) = 3$$

De la misma forma, la cantidad de operarios que obtuvo un puntaje perteneciente al intervalo 9 – 12, dentro del grupo experimentado, es:

$$20 - (1 + 3 + 6 + 2) = 8$$

Puntaje	Grupo Novato	Grupo Experimentado
17 – 20	2	1
13 – 16	**3**	3
9 – 12	8	**8**
5 – 8	5	6
1 – 4	2	2
Total	**20**	**20**

b) Cuartiles del grupo novato:

Completamos la tabla con las frecuencias acumuladas correspondientes al grupo.

El Q_1 es equivalente al C_{25}. El 25% de las observaciones es $\frac{k \cdot n}{100} = \frac{25 \cdot 20}{100} = 5$

Puntaje	Grupo Novato	n_i
17 – 20	2	20
13 – 16	3	18
9 – 12	8	15
5 – 8	**5**	**7**
1 – 4	2	2

El intervalo crítico es el intervalo 5 – 8.

$$C_{25} = 4{,}5 + \frac{4}{5} \cdot (5 - 2) \quad \Rightarrow \quad C_{25} = Q_1 = 6{,}9$$

El Q_2 es equivalente al C_{50}. El 50% de las observaciones es $\frac{k \cdot n}{100} = \frac{50 \cdot 20}{100} = 10$

Puntaje	Grupo Novato	n_i
17 – 20	2	20
13 – 16	3	18
9 – 12	**8**	**15**
5 – 8	5	7
1 – 4	2	2

El intervalo crítico es el intervalo 9 – 12.

$$C_{50} = 8{,}5 + \frac{4}{8} \cdot (10 - 7) \quad \Rightarrow \quad C_{50} = Q_2 = 10$$

El Q_3 es equivalente al C_{75}. El 75% de las observaciones es $\frac{k \cdot n}{100} = \frac{75 \cdot 20}{100} = 15$

El intervalo crítico es también el intervalo 9 – 12.

$$C_{75} = 8{,}5 + \frac{4}{8} \cdot (15 - 7) \quad \Rightarrow \quad C_{75} = Q_3 = 12{,}5$$

Grupo novato: Q_1 = 6,9; Q_2 = 10; Q3 = 12,5

c) En los dos grupos la mitad de los valores se ubican a izquierda y derecha de medidas “aproximadas” lo cual no da evidencias de que los años de experiencia de un obrero en su puesto favorecen el cumplimiento de las normas de seguridad y mantenimiento de las maquinarias.

Nota: ¿Cuándo dos medidas son “aproximadas”? ¿Por qué una diferencia de 0,5 en este caso puede considerarse “pequeña”? Nuevamente, como aclaramos en la respuesta del ejercicio 1, habría que tener en cuenta la variabilidad. Una variabilidad pequeña podría ser el producto del azar y no de la incidencia de los años de experiencia. Es necesario precisar cuándo una diferencia es “suficientemente pequeña” o “demasiado grande” para no ser explicada por el azar. Estas precisiones se obtienen mediante el empleo de técnicas estadísticas adecuadas, algunas de las cuales veremos hacia el final de este curso.

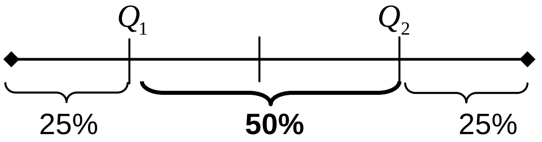
EJERCICIO 7

La distribución de frecuencias que se presenta muestra los resultados obtenidos en una prueba que mide el nivel de Escrupulosidad de los postulantes para un empleo de vigilancia privada. Los encargados de la selección decidieron optar por aquellos aspirantes que presentaran puntajes intermedios de Escrupulosidad. Esto significa que sólo quedará preseleccionado el 50% de los candidatos que obtienen las puntuaciones centrales.

X_i	n_i
13 - 15	25
10 - 12	23
7 - 9	14
4 - 6	6
1 - 3	2

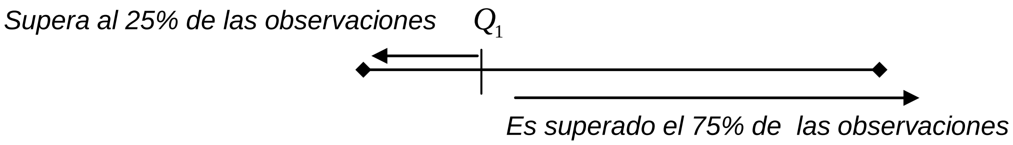
¿Qué puntuaciones como mínimo y como máximo será necesario obtener para quedar preseleccionado?

Las puntuaciones centrales que debemos buscar corresponden al intervalo limitado por los cuartiles primero y tercero: Q_1 y Q_2:



Puntajes intermedios de Escrupulosidad

Cálculo del primer cuartil:



El 25% de las observaciones es: $\frac{k \cdot n}{100} = \frac{25 \cdot 70}{100} = 17{,}5$

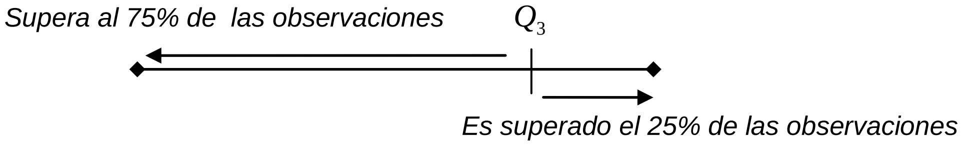
El intervalo crítico es el intervalo 7 – 9.

X_i	n_i	Frecuencia absoluta acumulada
13 – 15	25	60
10 – 12	23	45
7 – 9	**14**	**22**
4 – 6	6	8
1 – 3	2	2

Calculamos el centil 25:

$$C_{25} = 6{,}5 + \frac{3}{14} \cdot (17{,}5 - 8) \quad \Rightarrow \quad C_{25} = Q_1 = 8{,}54$$

Cálculo del tercer cuartil:



El 75% de las observaciones es: $\frac{k \cdot n}{100} = \frac{75 \cdot 70}{100} = 52{,}5$

El intervalo crítico es el intervalo 13 – 15.

X_i	n_i	Frecuencia absoluta acumulada
13 - 15	**25**	**60**
10 - 12	23	45
7 - 9	14	22
4 - 6	6	8
1 - 3	2	2

Calculamos el centil 75:

$$C_{75} = 12{,}5 + \frac{3}{25} \cdot (52{,}5 - 45) \quad \Rightarrow \quad C_{75} = Q_3 = 13{,}4$$

Puntaje mínimo: Q_1 = 8,54   Puntaje máximo: Q_3 = 13,4.

EJERCICIO 8

El director de un geriátrico decide analizar la cantidad de visitas semanales que recibe cada uno de los ancianos internados con el fin de tomar precauciones contra los síntomas depresivos que puede producir el sentimiento de abandono. Él considera que debe despreocuparse de aquellos ancianos que reciban más de 4 visitas, mientras que deberá evaluar psicológicamente a aquellos cuya cantidad de visitas no supere el decil 3. Los siguientes datos corresponden a la cantidad de visitas recibidas por los 30 ancianos del establecimiento.

0	5	2	1	7	6	5	0	2	3	5	1	2	2	5
7	6	1	3	2	4	1	4	0	2	5	6	3	4	1

a) Presente la información en una tabla de frecuencias absolutas (sin agrupar los datos en intervalos de clase).
b) ¿Qué porcentaje de casos no preocupan al director?
c) ¿Qué cantidad máxima de visitas debe tener el anciano para ser evaluado psicológicamente?
d) Si el director desea seleccionar a aquellos ancianos cuya cantidad de visitas ha superado el centil 80 o no ha superado al centil 20 para realizarles un test de personalidad ¿cuántos ancianos debe elegir?

a)

Cantidad de visitas	n_i
0	3
1	5
2	6
3	3
4	3
5	5
6	3
7	2

b) Del total de ancianos, 10 son los que reciben más de 4 visitas. Calculamos qué porcentaje representan sobre el total de 30 ancianos:

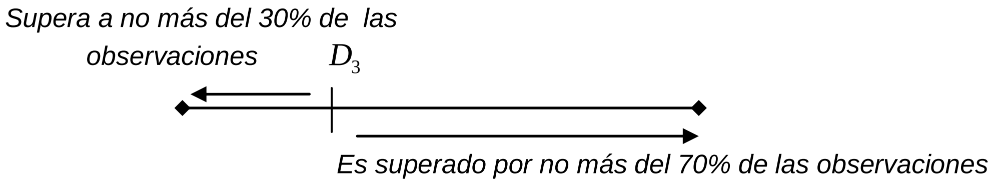
30 ancianos ——————— 100%
10 ancianos ——————— x

$$x = \frac{10 \cdot 100}{30} = 33{,}\widehat{3}\%$$

Se despreocupará del 33,33% de los casos.

c) El director evaluará psicológicamente a aquellos ancianos cuya cantidad de visitas no supere el decil 3. Se trata de una variable cuantitativa discreta. Para encontrar dicha puntuación debemos calcular las *frecuencias porcentuales acumuladas* en ambos sentidos.

Recordemos que el D_3 es la puntuación que supera a no más del 30% de las observaciones y es superada por no más del 70% de las mismas.



Cantidad de visitas	n_i	p_i	P_i%	f%	g%
0	3	0.100	10	10	100
1	5	0.167	16.7	**26.7**	90
2	**6**	**0.200**	**20**	**46.7**	**73.3**
3	3	0.100	10.0	56.7	**53.3**
4	3	0.100	10.0	66.7	43.3
5	5	0.167	16.7	83.3	33.3
6	3	0.100	10	93.3	16.7
7	2	0.067	6.7	100	6.7

Si observamos nuestra tabla de distribución de frecuencias, la puntuación que cumple ambas condiciones es 2, pues el 26,7% de los ancianos recibieron menos de 2 visitas y el 53,3% recibió más de 2.

Debe ser asistido psicológicamente todo anciano que tenga a lo sumo 2 visitas semanales.

d) El C_{20} es la puntuación que supera a no más del 20% de las observaciones y es superada por no más del 80% de las mismas; la puntuación que cumple con estas condiciones es 1, pues el 10% de los ancianos recibió menos de 2 visitas semanales y el 73,3% de ellos recibió más de 2 visitas semanales.

El C_{80} es la puntuación que supera a no más del 80% de las observaciones y es superada por no más del 20% de las mismas; la puntuación que cumple con estas condiciones es 5, pues el 66,7% de los ancianos recibió menos de 5 visitas semanales y el 16,7% de ellos recibió más de 5 visitas semanales.

El director debe seleccionar, por un lado, a aquellos ancianos cuya cantidad de visitas ha superado el C_{80}, es decir, a aquellos ancianos que han recibido una cantidad de visitas mayor a 5; por lo tanto, debe seleccionar **5** ancianos (3 ancianos que reciben 6 visitas semanales y 2 que reciben 7 visitas semanales). También debe seleccionar a aquéllos cuya cantidad de visitas no supere al C_{20}, esto es, 1 o menos cantidad de visitas semanales; si observamos la tabla debe seleccionar **5** ancianos que reciben 1 visita y **3** que no reciben ninguna.

Cantidad de visitas	n_i	p_i	P_i%	f%	g%
0	3	0.100	10	**10**	100
1	**5**	**0.167**	**16.7**	**26.7**	**90**
2	6	0.200	20	46.7	**73.3**
3	3	0.100	10.0	56.7	53.3
4	3	0.100	10.0	**66.7**	43.3
5	**5**	**0.167**	**16.7**	**83.3**	**33.3**
6	3	0.100	10	93.3	**16.7**
7	2	0.067	6.7	100	6.7

Serán seleccionados en total 13 ancianos.

MEDIDAS DE TENDENCIA CENTRAL

Las medidas de tendencia central son un tipo especial de medida de posición que da cuenta de la magnitud general de las observaciones.

Ellas son:

Media Aritmética: Es la suma de todos los valores observados dividida por el número de ellos.

Su cálculo tiene sentido a partir del nivel intervalar.

Mediana: Es el valor que supera y es superado por, como máximo, la mitad de las observaciones. Si la variable es continua, su cálculo se obtiene aplicando la fórmula de centiles, es decir, calculando el C_{50}. Si no lo es, acumulando las frecuencias porcentuales en ambos sentidos.

Su definición supone un ordenamiento de los valores, por lo tanto, **su cálculo tiene sentido a partir del nivel ordinal**.

Moda: Es el valor de la variable observado con más frecuencia.

Puede usarse en todos los niveles de medición.

EJERCICIO 9

Treinta ratas fueron separadas aleatoriamente en dos grupos iguales con la finalidad de comparar el efecto de la anestesia en las intervenciones quirúrgicas. El Grupo Tratamiento recibió un simulacro de operación que consistió en anestesiar al animal y efectuarle un corte inocuo, mientras que el Grupo Control no fue intervenido. Un día después se registró el tiempo que cada una necesitó para recorrer un laberinto. Esta tabla resume los resultados de ambos grupos:

Tiempo (seg.)	Grupo Tratamiento	Grupo Control
55 - 59	1	-
50 - 54	2	4
45 - 49	5	3
40 - 44	4	6
35 - 39	3	2

a) Calcule la media, la mediana y la moda aproximada[1] de ambas distribuciones.
b) ¿Observa diferencias en el rendimiento de las ratas entre las anestesiadas y las no anestesiadas? Conteste a partir del estadístico más adecuado.

a) ***Grupo Tratamiento***

Cálculo de la media:

Para hallar la media, se asume el supuesto de concentración en el punto medio del intervalo. Así, en el intervalo 35 – 39, se trata de sumar 3 valores iguales a 37, en el intervalo 40 – 44, sumar 4 valores iguales a 42, etc. Luego, se divide la suma de todos estos valores por el número de observaciones.

Completamos los datos con los puntos medios de cada intervalo, y el producto de cada uno de ellos por su frecuencia:

Tiempo (seg.)	Punto Medio (X_i)	n_i	$X_i \cdot n_i$
55 - 59	57	1	57
50 - 54	52	2	104
45 - 49	47	5	235
40 - 44	42	4	168
35 - 39	37	3	111
Total		**15**	**675**

$$\bar{x} = \frac{\sum X_i \cdot n_i}{n} \qquad \Rightarrow \qquad \bar{x} = \frac{675}{15} = 45$$

Media = **45 segundos**

Cálculo de la mediana:

El cálculo de la mediana en una distribución de frecuencias de una variable cuantitativa continua se obtiene aplicando la fórmula del C_{50}.

El 50% de los datos es: $\frac{k \cdot n}{100} = \frac{50 \cdot 15}{100} = 7{,}5$

El intervalo crítico es el intervalo 45 – 49.

Tiempo (seg.)	Punto Medio (X_i)	n_i	n_a
55 - 59	57	1	15
50 - 54	52	2	14
45 - 49	**47**	**5**	**12**
40 - 44	42	4	7
35 - 39	37	3	3

$$C_{50} = 44{,}5 + \frac{5}{5} \cdot (7{,}5 - 7) \quad \Rightarrow \quad C_{50} = 45$$

Mediana: **45 segundos**

---

[1] Existe una fórmula para la moda de datos agrupados en intervalos de clase que toma en cuenta la marca de clase del intervalo de mayor frecuencia y también la frecuencia de los dos intervalos adyacentes. Por simplicidad no utilizaremos esa fórmula y aproximaremos con el punto medio del intervalo modal.

Moda: **47 segundos**

Grupo Control

Cálculo de la media:

Procedemos de la misma forma que en el grupo anterior.

Tiempo (seg.)	Punto Medio (X_i)	n_i	$X_i \cdot n_i$
55 - 59	57	-	0
50 - 54	52	4	208
45 - 49	47	3	141
40 - 44	42	6	252
35 - 39	37	2	74
Total		**15**	**675**

$$\bar{x} = \frac{\sum X_i \cdot n_i}{n} \quad \Rightarrow \quad \bar{x} = \frac{675}{15} = \text{ } \textbf{45 segundos}$$

Cálculo de la mediana:

El 50% de los datos es: $\frac{k \cdot n}{100} = \frac{50 \cdot 15}{100} = 7,5$

El intervalo crítico es el intervalo 45 – 49.

Tiempo (seg.)	Punto Medio (X_i)	n_i	n_a
55 - 59	57	-	15
50 - 54	52	4	15
45 - 49	47	3	11
40 - 44	**42**	**6**	**8**
35 - 39	37	2	2

$$C_{50} = 39,5 + \frac{5}{6} \cdot (7,5 - 2) \quad \Rightarrow \quad C_{50} = 44,08 \quad \Rightarrow \quad \text{Me} = \textbf{44,08 segundos}$$

Moda: **42 segundos**

Grupo Tratamiento	Media = 45 seg	Mediana = 45 seg	Moda aprox= 47 seg
Grupo Control	Media = 45 seg	Mediana = 44,08 seg	Moda aprox= 42 seg

b) Elegimos la media por ser una variable cuantitativa que no ha tomado valores extremos no compensados que pudieran distorsionarla. Al ser igual en ambos grupos concluimos que el uso de anestesia en las operaciones parece no influir en el rendimiento.

EJERCICIO 10

Una vez tomados todos los tiempos del Ejercicio anterior los investigadores descubrieron que un error en el instrumento de medición agregó 10 segundos a cada una de las marcaciones. Indique cuáles son los valores corregidos de la media, la mediana y la moda.

Al sumar una constante a un conjunto de puntuaciones, la media aritmética quedará aumentada también en esa misma constante. Por lo tanto, para obtener el valor corregido de la misma, bastará con restarle 10 segundos al resultado obtenido en el ejercicio anterior, es decir, restamos 10 segundos de 45 segundos en ambos grupos.

En el caso de la mediana procedemos de la misma forma, restamos 10 segundos al valor obtenido anteriormente en cada grupo. En cada intervalo la corrección implica restar 10 segundo a ambos límites, con lo cual, en la fórmula del C_{50} sólo se modifica el límite exacto inferior del intervalo crítico, que se verá disminuido en esa cantidad, el resto de los valores de la fórmula no varía, es por eso que no es necesario hacer nuevamente los cálculos.

En el Grupo Tratamiento debemos cambiar 44,5 por 34,5:

$$C_{50} = \mathbf{44{,}5} + \frac{5}{5}\cdot(7{,}5-7) \quad \Rightarrow \quad C_{50} = \mathbf{34{,}5} + \frac{5}{5}\cdot(7{,}5-7)$$

En el Grupo Control debemos cambiar 29,5 por 19,5:

$$C_{50} = \mathbf{29{,}5} + \frac{5}{6}\cdot(7{,}5-2) \quad \Rightarrow \quad C_{50} = \mathbf{19{,}5} + \frac{5}{6}\cdot(7{,}5-2)$$

En el caso de la moda, dicho valor es el que tiene mayor frecuencia absoluta, con lo cual, también, sólo restamos 10 segundos a los valores obtenidos anteriormente.

Grupo Tratamiento	Media = 35 seg	Mediana = 35 seg	Moda = 37 seg
Grupo Control	Media = 35 seg	Mediana = 34,08 seg	Moda = 32 seg

EJERCICIO 11

Se ha realizado un estudio a fin de conocer los niveles de estrés laboral percibidos por los habitantes de una ciudad argentina y se ha pensado en la profesión como una posible fuente sistemática de variación. Aquí reproducimos los puntajes obtenidos en la prueba que mide el estrés laboral para cada una de las profesiones:

Grupo de maestros	45 – 32 – 29 – 49 – 40 – 44 – 33 – 46 – 45 – 48
Grupo de bomberos	41 – 36 – 31 – 36 – 39 – 42 – 35 – 36 – 40
Grupo de abogados	38 – 25 – 46 – 41 – 37 – 50 – 43
Grupo de médicos	43 – 37 – 50 – 53 – 40 – 35 – 56 – 49 – 55 – 45

Calcule la mediana de cada uno de los grupos y responda, en función de este estadístico, cuál parece ser el grupo más afectado por el estrés.

Disponemos de conjuntos de datos sin agrupar. Para calcular la mediana ordenamos las puntuaciones en sentido creciente. Nos encontramos frente a dos casos generales: si la cantidad de valores es *impar*, la mediana será la puntuación que ocupe el lugar central del ordenamiento; si se trata de una cantidad *par* de valores, la mediana estará formada por los dos valores centrales de dicho ordenamiento (si su cálculo lo permite, se calcula la media aritmética de ambos valores).

Grupo de maestros: 29 – 32 - 33 – 40 – **44** – **45** – 45 – 46 – 48 - 49

$$\text{Me} = \frac{44+45}{2} = 44{,}5$$

Grupo de bomberos: 31 – 35 - 36 – 36 – **36** – 39 – 40 – 41 – 42

Me = 36

Grupo de abogados: 25 – 37 - 38 – **41** – 43 – 46 – 50

Me = 41

Grupo de maestros: 35 – 37 - 40 – 43 – **45** – **49** – 50 – 53 – 55 - 56

$$\text{Me} = \frac{45+49}{2} = 47$$

Maestros = 45,5	Bomberos = 36	Abogados = 41	Médicos = 47

En función de los datos analizados, el grupo de médicos percibe un mayor nivel de estrés laboral.

EJERCICIO 12

Con el objetivo de analizar la influencia del contexto en la resolución de problemas de matemática un grupo de investigadores evaluó el desempeño de 90 niños que ayudan a sus padres en la atención de los puestos de una feria. La experiencia constaba de: 1) una prueba informal realizada en la feria en donde un entrevistador proponía una situación problemática de compra-venta de atención al público y 2) una prueba formal que consistía en la resolución de operaciones aritméticas equivalentes a las de la prueba informal pero sin la situación de compra-venta de atención al público. Se eligieron 45 niños al azar entre los 90 para administrarle la prueba informal y al resto se les administró la prueba formal.

Cantidad de problemas resueltos correctamente	Prueba Informal	Prueba Formal
54 – 62	7	3
45 – 53	15	5
36 – 44	9	3
27 – 35	7	9
18 – 26	6	14
9 – 17	0	9
0 – 8	1	2

a) Calcule media, mediana y moda aproximada de ambos grupos.
b) Extraiga conclusiones a partir de los resultados obtenidos.

Nota: El contenido de este ejercicio está inspirado en una investigación de Caraher, Carraher y Schlieman (1998). Remitimos al alumno interesado en los hallazgos y conclusiones de este estudio a consultar la publicación citada en la bibliografía.

I. Cálculo de la media, mediana y moda para el grupo de niños a los que se administro la prueba informal.

Para el cálculo de la media comenzamos por hallar el punto medio de cada intervalo y los productos de cada uno de ellos por su frecuencia absoluta.

Cantidad de problemas resueltos correctamente	Punto Medio (X_i)	Prueba Informal	$X_i \cdot n_i$
54 – 62	58	7	406
45 – 53	49	15	735
36 – 44	40	9	360
27 – 35	31	7	217
18 – 26	22	6	132
9 – 17	13	0	0
0 – 8	4	1	4
Total		**45**	**1854**

$$\bar{x} = \frac{1854}{45} = \mathbf{41{,}2}$$

Cálculo de la mediana:

El 50% de las observaciones es: $\frac{k \cdot n}{100} = \frac{50 \cdot 45}{100} = 22{,}5$

El intervalo crítico es el intervalo 36 – 44.

Cantidad de problemas resueltos correctamente	Punto Medio (X_i)	Prueba Informal	n_a
54 – 62	58	7	45
45 – 53	49	15	38
36 – 44	**40**	**9**	**23**
27 – 35	31	7	14
18 – 26	22	6	7
9 – 17	13	0	1
0 – 8	4	1	1

$$C_{50} = 35{,}5 + \frac{9}{9} \cdot (22{,}5 - 14) \quad \Rightarrow \quad C_{50} = 44 \quad \Rightarrow \quad \text{Me} = \mathbf{44}.$$

La moda es **49**.

II. Cálculo de la media, mediana y moda para el grupo de niños a los que se administro la prueba formal.

Media:

Cantidad de problemas resueltos correctamente	Punto Medio (X_i)	Prueba Formal	$X_i \cdot n_i$
54 – 62	58	3	174
45 – 53	49	5	245
36 – 44	40	3	120
27 – 35	31	9	279
18 – 26	22	14	308
9 – 17	13	9	117
0 – 8	4	2	8
Total		**45**	**1251**

$$\bar{x} = \frac{1251}{45} = \mathbf{27{,}8}$$

Mediana:

El 50% de las observaciones es: $\frac{k \cdot n}{100} = \frac{50 \cdot 45}{100} = 22{,}5$

El intervalo crítico es el intervalo 18 – 26.

Cantidad de problemas resueltos correctamente	Punto Medio (X_i)	Prueba Formal	n_a
54 – 62	58	3	45
45 – 53	49	5	42
36 – 44	40	3	37
27 – 35	31	9	34
18 – 26	**22**	**14**	**25**
9 – 17	13	9	11
0 – 8	4	2	2

$C_{50} = 17{,}5 + \frac{9}{14} \cdot (22{,}5 - 11) \Rightarrow C_{50} = 24{,}9 \Rightarrow$ Me = **24,9**.

La moda es **22**.

a)

Prueba informal	Media = 41,2	Mediana = 44	Moda = 49
Prueba Formal	Media = 27,8	Mediana = 24,9	Moda = 22

b) Las medidas de tendencia central resultaron mucho mayores para la prueba informal que para la formal; es decir, la cantidad de problemas bien resueltos resultó en general mayor cuando éstos se plantearon en contexto real.

EJERCICIO 13

Un investigador supone que la cantidad de público presente en un auditorio provoca reacciones psicofisiológicas que afectan al orador. Para probar esto, se trabajó con 26 conferencistas que debían pronunciar un discurso de igual complejidad y que fueron repartidos aleatoriamente en dos grupos: uno con el auditorio completo de público y otro con el auditorio con escaso público. Mientras tanto, un equipo registraba las pulsaciones por minuto ofreciendo, al final de la experiencia, un valor promedio de las mismas para cada uno de los conferencistas. A continuación se reproducen los datos:

Cantidad de pulsaciones por minuto	Auditorio completo de público	Auditorio con escaso público
101 – 109	2	1
92 – 100	6	1
83 – 91	5	3
74 – 82	2	4
65 – 73	1	1

a) Calcule la cantidad de pulsaciones promedio por minuto de cada grupo.
b) Extraiga conclusiones a partir de los resultados de la experiencia.

I. Cálculo de la cantidad de pulsaciones promedio por minuto para el grupo de conferencistas que pronunció el discurso con auditorio completo de público.

Cantidad de pulsaciones por minuto	Punto Medio (X_i)	Auditorio completo de público	$X_i \cdot n_i$
101 – 109	105	2	210
92 – 100	96	6	576
83 – 91	87	5	435
74 – 82	78	2	156
65 – 73	69	1	69
Total		**16**	**1446**

$$\bar{x} = \frac{1446}{16} = 90{,}375$$

II. Cálculo de la cantidad de pulsaciones promedio por minuto para el grupo de conferencistas que pronunció el discurso con auditorio con escaso público.

Cantidad de pulsaciones por minuto	Punto Medio (X_i)	Auditorio con escaso público	$X_i \cdot n_i$
101 – 109	105	1	105
92 – 100	96	1	96
83 – 91	87	3	261
74 – 82	78	4	312
65 – 73	69	1	69
Total		**10**	**843**

$$\bar{x} = \frac{843}{10} = 84{,}3$$

a)

Alta densidad de público	Media = 90,375 pulsaciones/min	Baja densidad de público	Media = 84,3 pulsaciones/min

b) La densidad de público parece ser una fuente sistemática de variación, siendo en promedio mayor el número de pulsaciones por minuto ante alta densidad de público.

EJERCICIO 14

En cierta comisión de Estadística la media de las calificaciones del primer parcial fue 6 para las mujeres y 5 para los varones. Calcule la media de todas las calificaciones de dicha comisión sabiendo que corresponden a 20 mujeres y 10 varones.

Se trata de calcular la media total de un grupo de puntuaciones, cuando se conocen los tamaños y medias de dos subgrupos hechos a partir del grupo total, las 20 mujeres y los 10 varones. Estos dos grupos son mutuamente exclusivos y exhaustivos. En este caso, la media total puede obtenerse ponderando las medias parciales a partir de los tamaños de cada subgrupo en el que han sido calculadas (propiedad de la media).

$$\bar{x} = \frac{n_1 \cdot \bar{x}_1 + n_2 \cdot \bar{x}_2}{n_1 + n_2} \qquad \Rightarrow \qquad \bar{x} = \frac{20 \cdot 6 + 10 \cdot 5}{20 + 10} = \frac{170}{30} = 5{,}67$$

Media = **5,67**

EJERCICIO 15

En una empresa el personal está categorizado según se indica a continuación.

Categorías	Salario	Nº de empleados
1	$ 1.300	12
2	$ 1.500	31
3	$ 1.700	27
4	$ 2.200	7
5	$ 3.000	4
6	$ 5.000	2
7	$ 7.000	3

a) Calcule las medidas de tendencia central adecuadas para la variable "Categoría".
b) Calcule las medidas de tendencia central adecuadas para la variable "Salario".
c) Si quisiera representar la escala salarial de la empresa cuál de las medidas calculadas en b) utilizaría en caso de ser:

  i. Sindicalista.
  ii. Directivo.
  iii. Economista objetivo.

a) Cálculo de las medidas de tendencia central para la variable "Categoría"

Se trata de una variable cuasicuantitativa. Sólo tiene sentido calcular la moda y la mediana.

La categoría modal es 2, es el valor con mayor frecuencia absoluta.

Para encontrar la mediana debemos calcular las frecuencias porcentuales acumuladas en ambos sentidos.

Categorías	Salario	Nº de empleados	p_i	$f_a\%$	$g_a\%$
1	$ 1.300	12	13.95	**13,95**	100.00
2	**$ 1.500**	31	36.05	**50.00**	86.05
3	**$ 1.700**	27	31.40	81.40	50.00
4	$ 2.200	7	8.14	89.53	**18.60**
5	$ 3.000	4	4.65	94.19	10.46
6	$ 5.000	2	2.32	96.51	5.81
7	$ 7.000	3	3.49	100.00	3.49

Hay dos valores que cumplen la condición de superar a no más del 50% de las observaciones y ser superado por no más del 50% de las mismas:

- La puntuación 2 supera al 13,95% de las observaciones y es superado por el 50% de las observaciones.
- La puntuación 3 supera al 50% de las observaciones y es superada por el 18,60% de las mismas.

En este caso no tiene sentido calcular el promedio de los dos valores de la variable ya que representan categorías a las que pertenecen empleados de la empresa, es una variable cuasicuantitativa. Decimos que la mediana es, entonces, el par (2,3).

La categoría modal es 2 y la categoría mediana: es el par (2,3)

a) Cálculo de las medidas de tendencia central para la variable "Salario"

La variable "Salario" es cuantitativa continua. Tiene sentido el cálculo de las tres medidas de tendencia central.

Salario modal = $1.500

Los valores que cumplen con la condición de superar a no más del 50% de las observaciones y ser superado por no más del 50% de las mismas son, en este caso, también los que corresponden a la segunda y tercera fila, $ 1.500 y $ 1.700. La mediana es el promedio de ambos valores:

$$\text{Me} = \frac{1.500+1.700}{2} = \frac{3.200}{2} = \$1.600$$

Cálculo de la media:

Para el cálculo de la media debemos hallar el promedio de todas las puntuaciones observadas. En este caso, los datos están nuevamente agrupados en una tabla de distribución de frecuencias. Completamos la tabla con el producto de cada puntuación por su frecuencia absoluta:

Categorías	Salario	Nº de empleados	$X_i \cdot n_i$
1	$ 1.300	12	15.600
2	$ 1.500	31	46.500
3	$ 1.700	27	45.900
4	$ 2.200	7	15.400
5	$ 3.000	4	12.000
6	$ 5.000	2	10.000
7	$ 7.000	3	21.000
TOTAL		86	**166.400**

$$\bar{x} = \frac{166.400}{86} = 1934{,}88$$

Salario modal = $1.500, Salario mediano = $1.600 (lo tomamos como representante del intervalo (1.500,1.700)). Salario medio = $1.934,88

c) i) La moda. ii) La media. iii) La mediana.

Un sindicalista utilizaría la moda para representar la escala salarial de la empresa, ya que su interés es defender el salario de los trabajadores; un empresario utilizaría la media, pues de esa forma pretendería justificar que el promedio del salario de sus empleados es alto, aunque el mismo esté distorsionado por los pocos salarios altos presentes entre las observaciones. En cambio un economista objetivo, usaría la mediana, valor que, en este caso, mejor representa el salario medio de un empleado de esta empresa.

MEDIDAS DE DISPERSIÓN

Otra propiedad importante de los conjuntos de datos es la que se refiere al grado en que estos se parecen o se diferencian entre sí, ya que dos conjuntos de datos pueden tener, por ejemplo, la misma media, y ser muy diferentes entre sí. Esta propiedad recibe el nombre de **variabilidad**, **dispersión** u **homogeneidad**. Se trata, ahora, de medir el grado de variación que hay en un conjunto de datos; en otras palabras, determinar cómo se concentran esos datos en torno al valor central.

EJERCICIO 16

Sobre la base de la información brindada en este cuadro responda las siguientes preguntas (los datos no son reales):

¿Qué diario compra habitualmente?				
Barrio	Clarín	La Nación	Página 12	Ninguno
Belgrano	39 %	51 %	5 %	5 %
Flores	40 %	20 %	20 %	20 %
Mataderos	47 %	16 %	11 %	26 %

a) ¿Cuál es el diario comprado más frecuentemente en cada barrio?
b) ¿Con qué medida compararía las distribuciones de los barrios con respecto a su variabilidad en el consumo de diarios?
c) ¿Qué distribución presenta mayor entropía: la del barrio de Belgrano o la del barrio de Flores?

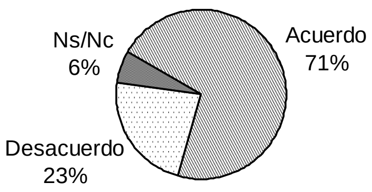
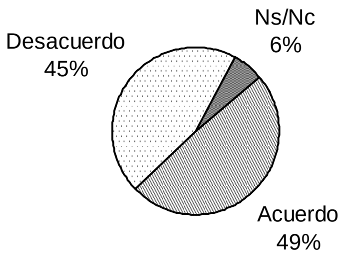
a) La información brindada en el cuadro muestra los porcentajes de personas que en cada barrio leen cada uno de los tres diarios mencionados o ninguno. Se trata de una variable cualitativa, con lo cual la medida de tendencia central que representa la magnitud general de los mismos (mayor concentración de observaciones) es la moda. Las modas son: “La Nación” para el barrio de Belgrano, el diario “Clarín” para las distribuciones de Flores y Mataderos.

b) Las variables cualitativas se miden a nivel nominal. En este nivel de medición la medida que nos permite evaluar y comparar la dispersión de dos o más conjuntos de datos es la ***entropía***. Definimos la entropía como el grado de incertidumbre de la clase a la que pertenece un dato.

c) El barrio de Flores. En el barrio de Belgrano la opinión está más concentrada en una modalidad, “La Nación”; es por eso que la entropía es menor que en el barrio de Flores, donde las opiniones están “más repartidas” entre las distintas modalidades.

EJERCICIO 17

<table>
<tr><td colspan="2">¿Está de acuerdo con la despenalización de la marihuana?</td><td rowspan="3">Ficha técnica
(Datos ficticios)

Encuesta realizada en la página web de un periódico porteño por 420 sujetos; 70% menores de 25 años y 30% mayores de 25, en el primer semestre de 2009.</td></tr>
<tr><td>Menores de 25</td><td>Mayores de 25</td></tr>
<tr><td> </td><td> </td></tr>
</table>

Teniendo en cuenta la información gráfica, responda lo siguiente:

a) ¿Cuál es la opinión más frecuente en cada grupo etáreo?
b) ¿En cuál de los grupos la opinión de la población está mejor representada por la moda?
c) ¿Qué distribución de opiniones presenta mayor entropía?
d) Indique con qué otro tipo de gráfico podría representar los datos si quisiera comparar los diferentes grupos etáreos.

a) "Acuerdo" en ambos casos.

b) En el caso de los menores de 25, pues más de la mitad de la población tiene esa opinión. Hay más concentración de opiniones en la moda que para los mayores. Cuando el número de clases es pequeño como en este caso, una gran concentración en la moda va acompañada de una menor entropía (queda poco por repartir entre las clases restantes).

c) La correspondiente a los mayores de 25.

d) Diagrama de rectángulos adyacentes.

EJERCICIO 18

Un investigador desea estudiar si el hecho de realizar una actividad o de permanecer ocioso durante un determinado lapso influye en la estimación subjetiva del tiempo transcurrido en dicho período. Efectuó una prueba piloto con 10 sujetos, la mitad permaneció inactivo (grupo inactivo) durante 25 minutos, la otra mitad realizó una tarea (grupo activo) que consistió en tomar nota del dictado de un texto. Ninguno de los dos grupos tenía la ayuda de un reloj. Al finalizar los 25 minutos los 10 sujetos debieron estimar el lapso transcurrido y resultó:

Grupo Inactivo	30	23	28	19	26
Grupo Activo	23	24	26	18	29

a) Calcule la amplitud total, la media y el desvío típico de ambos grupos.
b) ¿Qué indicios dan estos resultados respecto de la hipótesis de la investigación?
c) ¿Qué grupo ha resultado más variable en su estimación del tiempo?

a) La estimación subjetiva del tiempo transcurrido en ese período de tiempo es una variable cuantitativa continua. Ya hemos definido, en la práctica II, la **amplitud total** como la diferencia entre el mayor y el menor de los valores observados en la muestra.

$$A_T = X_{máx} - X_{mín}$$

Otra posible forma de medir el grado de dispersión de un conjunto de puntuaciones se basa en la idea de trabajar con las distancias desde cada valor a un valor central, en este caso, la media, y luego obtener su promedio. Para ello debemos calcular en primer lugar las puntuaciones diferenciales. Su suma siempre es igual a cero, es por eso que se procede de dos formas diferentes; una es tomar esas diferencias en valor absoluto, la otra es elevar cada una de ellas al cuadrado.

Definimos así, dos nuevas medidas de variación:

Desvío Medio: Es el promedio de las distancias que existe entre cada valor y la media.

Datos sin agrupar: $DM = \frac{\sum |x_i - \bar{x}|}{n}$

Datos agrupados: $DM = \frac{\sum |x_i - \bar{x}| \cdot n_i}{n}$

Varianza: Es el promedio del cuadrado de las distancias a la media. Su raíz cuadrada es el **desvío típico**.

Datos sin agrupar: $S^2 = \frac{\sum (x_i - \bar{x})^2}{n}$

Datos agrupados: $S^2 = \frac{\sum (x_i - \bar{x})^2 \cdot n_i}{n}$

Calculamos las medidas de variabilidad para cada grupo:

Grupo Inactivo:

Amplitud Total: $A_T =$ 30 – 19 = **11**

Media: $\bar{x} = \frac{30 + 23 + 28 + 19 + 26}{5} =$ **25,2**

Varianza:

$$S^2 = \frac{(30-25{,}2)^2 + (23-25{,}2)^2 + (28-25{,}2)^2 + (19-25{,}2)^2 + (26-25{,}2)^2}{5} = 14{,}96$$

Desvío Típico: $S = \sqrt{14{,}96} =$ **3,87**

Grupo Activo:

$A_T =$ 29 – 18 = **11**

Media: $\bar{x} = \frac{23 + 24 + 26 + 18 + 29}{5} =$ **24**

Varianza: $S^2 = \frac{(23-24)^2 + (24-24)^2 + (26-24)^2 + (18-24)^2 + (29-24)^2}{5} = 13{,}2$

Desvío Típico: $S = \sqrt{13{,}2} =$ **3,63**

Grupo Inactivo	AT = 11	Media = 25,2	S = 3,87
Grupo Activo	AT = 11	Media = 24	S = 3,63

b) Los resultados van en la dirección de la hipótesis de que el nivel de actividad influye en la estimación subjetiva del tiempo transcurrido, en el sentido de que los inactivos tienden a percibirlo como mayor. Sin embargo para inferir esto más allá de las muestra en cuestión hay que completar el análisis con un contraste de hipótesis.

c) En esta muestra el grupo inactivo ha resultado levemente más variable tanto en términos absolutos (desviación estándar) como relativos a la media; ya que los coeficientes de variación son 15,36 y 15,12 respectivamente. Para determinar si esta diferencia aparentemente leve es estadísticamente significativa también se debería llevar a cabo el correspondiente contraste de hipótesis.

EJERCICIO 19

¿Cómo habrían resultado los resúmenes estadísticos calculados en el Ejercicio 18 si…

a) todos los datos hubieran estado incrementados en 2 minutos;
b) todos los tiempos hubieran sido percibidos en un 50% más largos?

a) Siendo la AT la distancia entre el mayor y el menor valor observados, incrementar ambos valores en 2 no altera el resultado de dicha medida:

Grupo Inactivo:

$A_T =$ 30 – 19 = 11

$A_T =$ (30 **+ 2**) – (19 **+ 2**) = 30 **+ 2** – 19 **– 2** = 30 – 19 = **11**

Grupo Activo:

$A_T =$ 29 – 18 = 11

$A_T =$ (29 **+ 2**) – (18 **+ 2**) = 29 **+ 2** – 18 **– 2** = 29 – 18 = **11**

La transformación tampoco altera los resultados de la desviación típica ya que si sumamos una constante a un conjunto de puntuaciones, por una de las propiedades de la varianza, ésta no varía, por ende, tampoco variará S.

Con respecto a la media aritmética, al sumar una constante a un conjunto de puntuaciones, la media quedará aumentada en esa misma constante. En este caso, se ha sumado 2 a cada puntuación, con lo que la media del nuevo conjunto de puntuaciones en cada grupo será:

Grupo Inactivo:

Media: $\bar{x}$ = 25,2 + **2** = **27,2**

Grupo Activo:

Media: $\bar{x}$ = 24 + **2** = **26**

Grupo Inactivo	AT = 11	Media = 27,2	S = 3,87
Grupo Activo	AT = 11	Media = 26	S = 3,63

b) Incrementar los valores en un 50% es lo mismo que aplicar la transformación Y=1,5X; por lo que los resúmenes estadísticos resultarían:

La amplitud total quedará multiplicada por dicha constante:

Grupo Inactivo:

$A_T =$ 30 – 19 = 11

$A_T =$ **1,5** · 30 – **1,5** · 19 = **1,5** · (30 – 19) = 1,5 · 11 = **16,5**

Grupo Activo:

$A_T =$ 29 – 18 = 11

$A_T =$ **1,5** · 29 – **1,5** · 18 = **1,5** · (29 – 18) = **1,5** · 11 = **16,5**

Teniendo en cuenta las propiedades de la desviación típica y de la media aritmética, la desviación típica del nuevo conjunto de puntuaciones será igual a la desviación típica del conjunto original multiplicada por el valor absoluto de la constante, es decir, por 1,5. La media aritmética, resultará igual a la del conjunto original multiplicada por la constante, 1,5.

Grupo Inactivo:

Media: $\bar{x}$ = 25,2 · **1,5** = **37,8**

Desvío Típico: S = 3,87 · **1,5** = **5,805**

Grupo Activo:

Media: $\bar{x}$ 24 · **1,5** = **36**

Desvío Típico: S = 3,63 · **1,5** = **5,445**

Grupo Inactivo	AT = 16,5	Media = 37,8	S = 5,805
Grupo Activo	AT = 16,5	Media = 36	S = 5,445

EJERCICIO 20

Los 28 docentes de una asignatura fueron divididos aleatoriamente en dos grupos para calificar un parcial. Mientras que el Grupo 1 no poseía referencia alguna sobre la calificación obtenida por el alumno en el primer parcial, el Grupo 2 sí la poseía.

Calificación	Grupo 1	Grupo 2
10	1	1
9	2	5
8	3	4
7	4	3
6	2	1
5	2	0

a) Calcule la media, la varianza y el desvío típico de ambos grupos.
b) Si a las calificaciones originales se las multiplica por 10, ¿cómo se modifican los resultados de la media, la varianza y el desvío típico?

a) ***Grupo 1***

Cálculo de la media:

Completamos la tabla con el producto de cada puntuación por su frecuencia absoluta:

Calificación	Grupo 1	$X_i \cdot n_i$
10	1	10
9	2	18
8	3	24
7	4	28
6	2	12
5	2	10
Total	**14**	**102**

$$\bar{x} = \frac{102}{14} = 7{,}29$$

Cálculo de la varianza:

Para calcular la varianza debemos obtener para cada puntuación el cuadrado de la diferencia entre dicha puntuación y la media y, luego, calcular su promedio.

Calificación	Grupo 1	$\left(x_i - \bar{x}\right)^2 \cdot n_i$
10	1	7.3441
9	2	5.8482
8	3	1.5123
7	4	0.3364
6	2	3.3282
5	2	10.4882
Total	**14**	**28,8574**

$$S^2 = \frac{28{,}8574}{14} = 2{,}06 \quad \Longrightarrow \quad S = \sqrt{2{,}06} = 1{,}44$$

Grupo 2

Cálculo de la media, varianza y desvío:

Calificación	Grupo 2	$X_i \cdot n_i$	$\left(x_i - \bar{x}\right)^2 \cdot n_i$
10	1	10	3,4596
9	5	45	3,698
8	4	32	0,0784
7	3	21	3,8988
6	1	6	4,5796
5	0	0	0
Total	**14**	**114**	**15,7144**

$$\bar{x} = \frac{114}{14} = 8{,}14$$

$$S^2 = \frac{15{,}7144}{14} = 1{,}12 \qquad \Rightarrow \qquad S = \sqrt{1{,}12} = 1{,}06$$

b) Aplicando las propiedades correspondientes a cada índice, la media de cada grupo queda multiplicada por la constante, 10; la desviación típica, por el valor absoluto de la constante, 10, y la varianza por el cuadrado de la constante, 100.

	Media		Varianza		Desviación Estándar	
	a)	b)	a)	b)	a)	b)
Grupo 1	7,29	72,9	2,06	206	1,44	14,4
Grupo 2	8,14	81,4	1,12	112	1,06	10,6

EJERCICIO 21

Para medir el tiempo reacción ante ciertos estímulos visuales se dispone del test A y para estímulos sonoros del test B. En la tabla siguiente se muestran los tiempos de reacción de Karina a 10 estímulos de cada tipo. ¿En cuál test ha tenido un desempeño más homogéneo?

. Tiempo (seg.)	Test A	Test B
33 – 39	3	-
26 – 32	3	1
19 – 25	2	2
12 – 18	1	3
5 – 11	1	4

Debemos comparar el desempeño de Karina en dos pruebas diferentes y ver en cuál de ellas ha sido más homogéneo. Como el tiempo de reacción corresponde a variables diferentes y, además, medibles ambas en el nivel de razón, usamos el **coeficiente de variación**. Este índice es una medida de variabilidad que expresa la desviación estándar relativa a la media en unidades de 100. Permite comparar la variabilidad de grupos con medias muy diferentes como, así también, la variabilidad de observaciones de distintas variables. Cuanto mayor es el CV menos representativa es la media del grupo de observaciones.

La fórmula que lo define es: $CV = \frac{S_x}{\bar{x}} \cdot 100$

En este caso, las puntuaciones obtenidas en cada test se encuentran agrupadas en intervalos, debemos obtener el punto medio de cada uno de ellos y luego para la obtención de la media y de la varianza, procedemos como en el ejercicio anterior.

. Tiempo (seg.)	Punto Medio (X_i)	Test A	Test B
33 – 39	36	3	-
26 – 32	29	3	1
19 – 25	22	2	2
12 – 18	15	1	3
5 – 11	8	1	4

Test A

.Tiempo (seg.)	Punto Medio (X_i)	Test A	$X_i \cdot n_i$	$\left(x_i - \bar{x}\right)^2 \cdot n_i$
33 – 39	36	3	108	288,12
26 – 32	29	3	87	23,52
19 – 25	22	2	44	35,28
12 – 18	15	1	15	125,44
5 – 11	8	1	8	331,24
Total		**10**	**262**	**803,6**

$$\bar{x} = \frac{262}{10} = 26,2$$

$$S^2 = \frac{803,6}{10} = 80,36 \quad \Rightarrow \quad S = \sqrt{80,36} = 8,9644$$

$$CV = \frac{8,9644}{26,2} \cdot 100 = \mathbf{34,22}$$

Test B

. Tiempo (seg.)	Punto Medio (X_i)	Test B	$X_i \cdot n_i$	$\left(x_i - \bar{x}\right)^2 \cdot n_i$
33 – 39	36	0	0	0
26 – 32	29	1	29	196
19 – 25	22	2	44	98
12 – 18	15	3	45	0
5 – 11	8	4	32	196
Total		**10**	**150**	**490**

$$\bar{x} = \frac{150}{10} = 15$$

$$S^2 = \frac{490}{10} = 49 \quad \Rightarrow \quad S = \sqrt{49} = 7$$

$$CV = \frac{7}{15} \cdot 100 = \mathbf{46,67}$$

Test A	Media = 26,2	S = 8,96	CV = 34,22
Test B	Media = 15	S = 7	CV = 46,67

Al comparar los coeficientes de variación se concluye que el desempeño de Karina ha sido más homogéneo en el test de estímulos visuales.

EJERCICIO 22

En un centro de psicopedagogía clínica 20 niños disléxicos, fueron divididos en dos grupos de 10 y recibieron el mismo tratamiento pero con diferentes terapeutas: un grupo con el terapeuta A y el otro con el terapeuta B. Al finalizar el tratamiento se les administró un test cuyos mayores puntajes están vinculados a mayor cantidad de palabras incorrectamente leídas. Los resultados obtenidos por ambos terapeutas son resumidos a continuación:

Terapeuta A	
Puntaje	ni
3	4
4	1
5	3
6	1
7	1

Terapeuta B
$x = 6$ Mdn = 5,5 Mo = 7 s = 1,44 CV= 24

Basándose sobre los resúmenes adecuados responda:

a) ¿Qué grupo obtuvo mejores resultados en el test?

b) ¿Puede afirmarse que el terapeuta A es más eficiente que el terapeuta B? Justifique.

a) Se trata de una variable cuantitativa. Si bien, es posible calcular todas las medidas de tendencia central, se prefiere, a menos que haya argumentos en contra, la media aritmética, ya que usa la toda la información de la que disponemos, es más estable y, además, es mejor estimador de su parámetro que la mediana y la moda.

Terapeuta A			
Puntaje	ni	$X_i \cdot n_i$	$\left(x_i - \bar{x}\right)^2 \cdot n_i$
3	4	12	7.84
4	1	4	0.16
5	3	15	1.08
6	1	6	2.56
7	1	7	6.76
Total	**10**	**44**	**18.4**

$$\bar{x} = \frac{44}{10} = 4{,}4$$

a) Obtuvo mejor resultado el grupo del terapeuta A ya que su media es 4,4; o sea menor que la del grupo del terapeuta B.

b) No puede afirmarse esto porque:

- No se sabe si ambos grupos tenían el mismo nivel de dificultad al momento de comenzar el tratamiento ya que no se hizo un pretest. Tampoco fueron aleatorizados como para suponer una distribución homogénea en los grupos.
- Aunque los dos grupos estuvieran en condiciones semejantes de partida habría que investigar si la diferencia entre las medias es suficientemente grande en relación a su variabilidad de muestra en muestra.

EJERCICIO 23

La estatura promedio de una muestra de 150 adultos es de 1,72 m con un desvío estándar de 0,07m.

a) Si se sabe que el promedio de peso de los sujetos de esta muestra es de 75 kg. con un desvío de 8 kg. ¿Es posible afirmar que los pesos son más variables que las estaturas porque les corresponde desvío estándar mayor?

b) Si se sabe que la estatura promedio de los hijos de los sujetos de la muestra es de 0,83 m con un desvío de 0,05 m. ¿Es posible afirmar que los padres varían más en la estatura que los hijos?

a) No, porque los desvíos estándar no son directamente comparables por estar en unidades diferentes. Sin embargo, puesto que ambas variables corresponden al nivel de razón, se puede comparar la variabilidad en relación a la media mediante el coeficiente de variación, que es adimensional. Éste es 4,1 para la estatura y 10,67 para el peso:

Estatura: $CV = \frac{0{,}07}{1{,}72} \cdot 100 = 4{,}1$

Peso: $CV = \frac{8}{75} \cdot 100 = 10{,}67$

Luego, son más variables los pesos que las estaturas, en relación a la media.

b) No es lo más adecuado comparar los desvíos en términos absolutos sino en relación a la media cuando el nivel de medición lo permite. Como la variable “estatura” corresponde al nivel de razón, podemos calcular el coeficiente de variación.

El CV para los hijos es: $CV = \frac{0{,}05}{0{,}83} \cdot 100 = 6$

Para los padres	CV = 4,1	La variabilidad relativa es mayor en los niños.
Para los hijos	CV = 6	

EJERCICIO 24

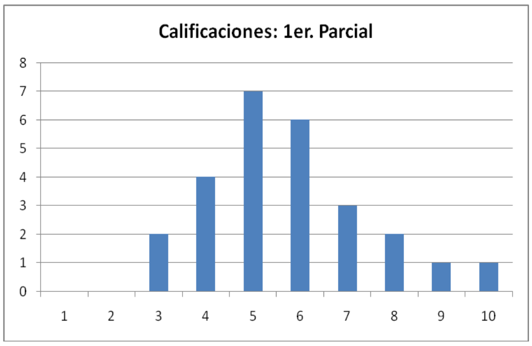
Los alumnos de una comisión de Escuela Francesa obtuvieron las calificaciones que se muestran en las siguientes tablas:

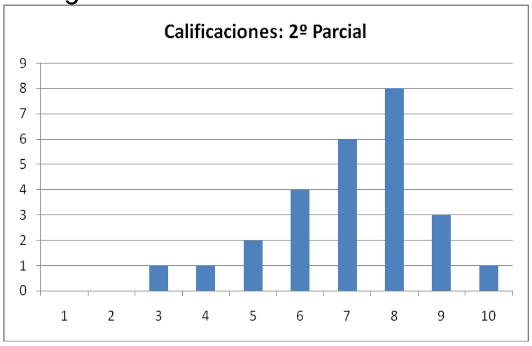
Calificaciones	Frecuencias	
	1º Parcial	2º Parcial
3	2	1
4	4	1
5	7	2
6	6	4
7	3	6
8	2	8
9	1	3
10	1	1

Grafique ambas distribuciones de modo que puedan ser comparadas y responda sin hacer cálculos:

a) ¿En qué parcial la comisión obtuvo mejor rendimiento?
b) ¿Qué parcial discriminó mejor a los alumnos de mejor rendimiento? ¿Y a los de peor rendimiento?
c) ¿Cómo debería ser la forma de la distribución para discriminar por igual a los alumnos de mejor y peor rendimiento?
d) Explique la vinculación entre el tipo de asimetría y el efecto discriminatorio de la prueba.

La calificación de cada parcial es una variable cuantitativa discreta. Para representar gráficamente ambas distribuciones utilizamos un diagrama de barras.





a) En el segundo parcial se obtuvo mejor rendimiento que en el primero pues, visualmente, se observa que la distribución es asimétrica negativa para el segundo parcial y asimétrica positiva para el primero.

b) El primer parcial discriminó más a los alumnos de mejor rendimiento, el segundo discriminó más a los de peor rendimiento.

c) Para discriminar a los de mejor y peor rendimiento por igual la forma de la distribución debería ser simétrica.

d) Los exámenes cuyos puntajes toman valores a lo largo de casi toda la escala de evaluación y tienen una distribución asimétrica negativa (abundan los aprobados y escasean los reprobados) son más fáciles que los que tienen una distribución asimétrica positiva (a la inversa, abundan los reprobados y son pocos los que aprobaron); por tanto los primeros discriminan mejor a los individuos de bajo rendimiento y los segundos a los de alto rendimiento.

EJERCICIO 25

En su artículo “Elementos de estadística al servicio de la evaluación del rendimiento”, la Lic. Floralba Cano de Becerra (1971) aconseja tomar pruebas que produzcan puntajes con asimetría positiva al inicio del curso y asimetría negativa en los exámenes finales. ¿Sobre qué criterio cree Ud. que se apoya esta sugerencia?

Si al inicio del curso los puntajes se distribuyen con asimetría positiva son pocos los alumnos que obtendrán los mayores puntajes y no tan pocos los que podrían sacar una nota baja. La intención sería incentivar al alumno desde el inicio del curso para que se esfuerce en no reprobar o en merecer notas excelentes. En cambio en un examen final interesa que las calificaciones bajas de reprobación sean infrecuentes y no tanto las excelentes; es decir que sea más difícil la reprobación que el premio de una nota alta;

para que si alguien reprueba se tenga la tranquilidad de que es porque realmente lo merece al destacarse por su mal rendimiento.

EJERCICIO 26

Con los datos del Ejercicio 7 de la Práctica I realice un análisis exploratorio a fin de determinar si la muestra da indicios de que el maestro y la modalidad de la clase son fuentes sistemáticas de variación para la rapidez y la comprensión lectora. (Use calculadora).

Resolveremos, a manera de ejemplo, el ejercicio para la variable R = Rapidez lectora y dejamos como ejercicio para el alumno el análisis para la variable C = Comprensión lectora.

Comenzaremos resumiendo la información. Dado que el nivel de medición lo permite, podemos representar los datos de las celdas por su media y resumir su variabilidad con el desvío estándar (usamos el que se obtiene de dividir por n). Codificaremos con CA y CN las modalidades de clases abreviadas y no abreviadas respectivamente, y con M1, M2 y M3 a los maestros 1, 2 y 3.

La tabla con los resúmenes queda así:

	CA		CN		Total	
	R	C	R	C	R	C
M1	$\bar{X} = 11$ s = 1,79	$\bar{X} =$ s =	$\bar{X} = 9,2$ s = 0,98	$\bar{X} =$ s =	$\bar{X} = 10,1$ s = 1,7	$\bar{X} =$ s =
M2	$\bar{X} = 13,4$ s = 1,36	$\bar{X} =$ s =	$\bar{X} = 10,2$ s = 1,17	$\bar{X} =$ s =	$\bar{X} = 11,8$ s = 2,04	$\bar{X} =$ s =
M3	$\bar{X} = 8$ s = 1,09	$\bar{X} =$ s =	$\bar{X} = 8,8$ s = 0,75	$\bar{X} =$ s =	$\bar{X} = 8,4$ s = 1,02	$\bar{X} =$ s =
Total	$\bar{X} = 10,8$ s = 2,64	$\bar{X} =$ s =	$\bar{X} = 9,4$ s = 1,14	$\bar{X} =$ s =	$\bar{X} = 10,1$ s = 2,15	$\bar{X} =$ s =

Antes de continuar responda las siguientes preguntas:

Pregunta 1:

¿Qué representan las medias y los desvíos del margen derecho de la tabla? ¿Y los de los últimos renglones?

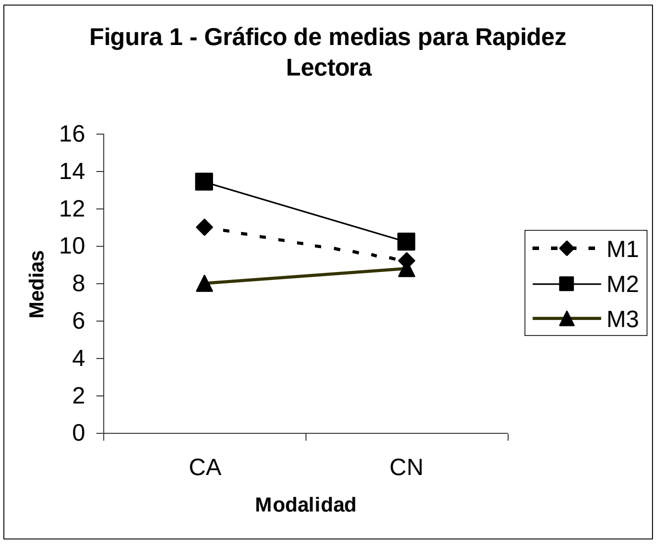
Pregunta 2:

Las medias marginales de las filas ¿coinciden con el promedio de las medias de sus respectivas filas? ¿Y las de las columnas? ¿Por qué?

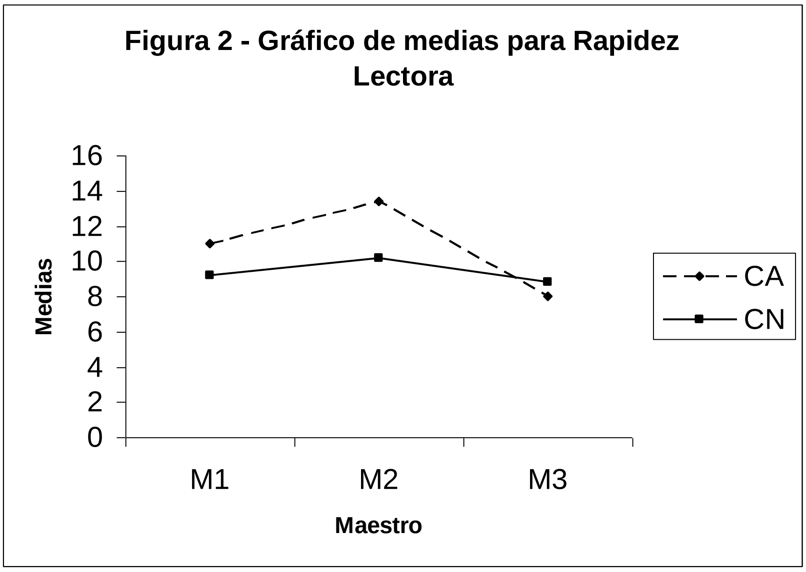
Pregunta 3:

Los desvíos estándar marginales de las filas ¿coinciden con las medias de los desvíos de las correspondientes filas? ¿Por qué?

Comparar los resultados entre maestros o entre modalidades de clase se reduce, en una simplificación del problema, a comparar las medias que los representan. Puede ser muy útil apoyarse en un gráfico de las mismas.



Los resultados son mejores para el maestro 2, luego para el maestro 1 y en último lugar para el maestro 3. Esto sucede en ambas modalidades de clase aunque las diferencias son notoriamente más pronunciadas en CA que en CN. Luego, según la muestra, parece haber un efecto del factor “maestro” dentro de cada modalidad. Recíprocamente se observa un efecto de la modalidad dentro de cada maestro, más pronunciado para los maestros 1 y 2 que para el maestro 3. Esto se aprecia por las pendientes de cada segmento (las primeras son decrecientes y la otra creciente) o bien graficando las medias de las modalidades para cada maestro a saber:



Este gráfico muestra más claramente que con clases abreviadas los maestros 1 y 2 obtuvieron mejores resultados que con las no abreviadas. El maestro 3 obtuvo una pequeña diferencia a favor de CN.

Dado que las diferencias entre los maestros no son iguales dentro de cada modalidad ni recíprocamente, se puede decir que sobre la variable R hay un efecto de la “combinación” de ambos factores: *maestro y modalidad de la clase*. Cuando esto ocurre se dice que hay “interacción” de los factores.

Como una manera de constatar si entendió estos conceptos le proponemos que responda a las siguientes preguntas: ¿Cómo sería la figura 1 si...

Pregunta 4: ... no hubiera efecto de la interacción pero si del maestro y de la modalidad de la clase?
Pregunta 5: ...hubiera efecto del factor maestro pero no de la modalidad de la clase?
Pregunta 6: ...no hubiera efecto del maestro pero sí de la modalidad?
Pregunta 7: ... no hubiera efecto de ninguno de ellos?

Resumiendo…

A partir de la inspección "ocular" de las medias podemos concluir sobre esta muestra que:

1) El maestro 2 obtiene mejores resultados siguiéndole el maestro 1 y luego el maestro 3, en ambas modalidades.
2) Los resultados entre los maestros difieren menos en las clases no abreviadas que en las abreviadas por lo que hay un efecto de la interacción maestro – modalidad.
3) Con las clases abreviadas se obtienen mejores resultados, excepto para el maestro 3.

Estas conclusiones ¿serán aplicables más allá de esta muestra? En otras palabras ¿se obtendrán resultados parecidos con otros estudiantes de similares características o, aun con los mismos, si éstos fueran asignados de otro modo a los distintos grupos? (Recordar que la asignación fue aleatoria). Si la respuesta fuera afirmativa podría concluirse, por ejemplo, que el maestro 2 se desempeña mejor que el 3 en la modalidad de clases abreviadas. La pregunta podría reformularse así: las diferencias observadas ¿son atribuibles a los factores considerados en el experimento o son atribuibles al azar; es decir, particulares de la muestra seleccionada? Para responder a esta pregunta hay criterios que proporcionan las técnicas de inferencia estadística algunas de las cuales, como anticipamos en el comentario del ejercicio 1, veremos al final de este curso. Por ahora sólo podemos tener un cierto indicio mirando las diferencias entre las medias y teniendo en cuenta el desvío estándar: cuanto mayor sea la diferencia y menor la variabilidad se podrá estar más seguro de que dicha diferencia es atribuible a una fuente sistemática. Por ejemplo: para CA se puede sospechar que efectivamente el M2 es mejor que el M3 porque hay una diferencia "amplia" de 5,4 unidades con un desvío "pequeño" de poco más de una unidad en ambos casos. La diferencia entre M1 y M2 bien podría deberse al azar puesto que es de 2,4 unidades con un desvío cercano a 2 para M1. Sin embargo, no hay indicios tan fuertes de que los maestros difieran en modalidad cuando sus clases no son abreviadas (CN); ya que las diferencias son mucho menos pronunciadas y los desvíos no tanto más pequeños, por tanto podrían deberse al azar y no al maestro.

Es claro que este comentario tiene ciertas imprecisiones ¿cuándo puede una diferencia de medias considerarse "amplia" o un desvío "pequeño"? Estas imprecisiones se superan con técnicas adecuadas desarrolladas por la teoría matemática de la inferencia estadística. Sin embargo es importante tener un "ojo entrenado" al abordar un problema aunque no se debería fiar sólo de él.

A continuación realizamos el análisis para la variable C = "Comprensión lectora". Mostramos la tabla con los resúmenes estadísticos correspondientes a dicha variable. Los cálculos fueron hechos con calculadora.

	CA		CN		Total	
	R	C	R	C	R	C
M1	$\bar{X} = 11$ s = 1,79	$\bar{X} = 21,2$ s =1,33	$\bar{X} = 9,2$ s = 0,98	$\bar{X} = 15,8$ s =1,17	$\bar{X} = 10,1$ s = 1,7	$\bar{X} = 18,5$ s =2,97
M2	$\bar{X} = 13,4$ s = 1,36	$\bar{X} = 24,8$ s =1,47	$\bar{X} = 10,2$ s = 1,17	$\bar{X} = 16,8$ s =1,17	$\bar{X} = 11,8$ s = 2,04	$\bar{X} = 20,8$ s =4,21
M3	$\bar{X} = 8$ s = 1,09	$\bar{X} = 17,2$ s =1,33	$\bar{X} = 8,8$ s = 0,75	$\bar{X} = 19$ s =1,67	$\bar{X} = 8,4$ s = 1,02	$\bar{X} = 18,1$ s =1,78
Total	$\bar{X} = 10,8$ s = 2,64	$\bar{X} = 21,1$ s =3,39	$\bar{X} = 9,4$ s = 1,14	$\bar{X} = 17,2$ s =1,90	$\bar{X} = 10,1$ s = 2,15	$\bar{X} = 19,1$ s =3,35

Pregunta 1: Las medias del margen derecho son las medias de los puntajes correspondientes a todos los alumnos de cada maestro, tomados globalmente sin discriminar la modalidad de clase. El desvío estándar corresponde al desvío de los puntajes de dichos alumnos para cada maestro. Análogamente, las medias y desvíos por columnas lo son de los alumnos correspondientes a cada modalidad sin tener en cuenta el maestro. La media y el desvío del último cuadro son las que se obtienen a partir de los 30 puntajes.

Pregunta 2: Las medias marginales pueden obtenerse promediando las medias por fila y por columna ya que cada celda corresponde a la misma cantidad de estudiantes: 5.

Pregunta 3: Los desvíos estándar no coinciden con los promedios de los desvíos por fila y por columna por razones de álgebra elemental; entre otras, la falta de distributividad de los cuadrados y de las raíces para la suma.

Pregunta 4: Los segmentos deberían ser paralelos no coincidentes y no horizontales.

Pregunta 5: Los segmentos deberían ser paralelos no coincidentes y horizontales.

Pregunta 6: Los segmentos deberían ser coincidentes no horizontales.

Pregunta 7: Los segmentos deberían ser coincidentes horizontales.

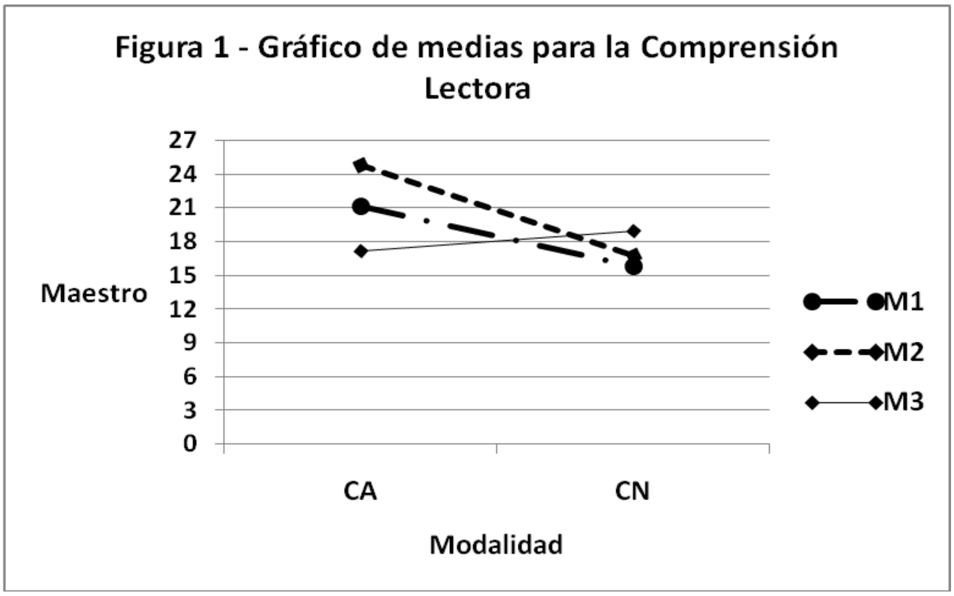
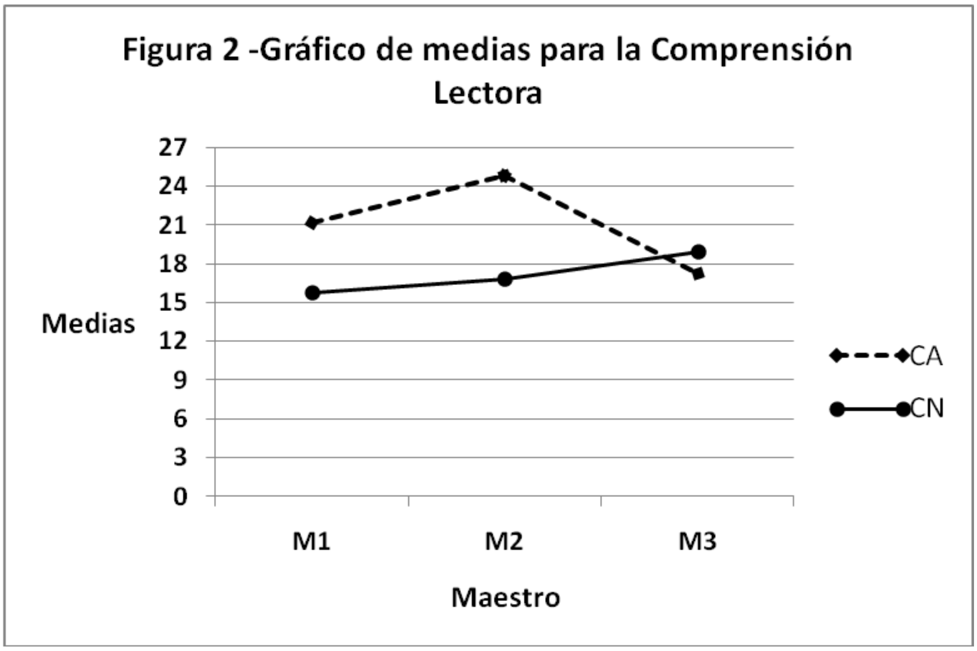


Figura 1 - Gráfico de medias para la Comprensión Lectora

En la modalidad CA, los resultados son mejores para el maestro 2, luego para el maestro 1 y en último lugar para el maestro 3, mientras que en la modalidad de CN los resultados son mejores para el maestro 3, luego para el maestro 2 por último para el M1; las diferencias son más pronunciadas en CA que en CN. Según la muestra, al igual que en el análisis de la variable Rapidez Lectora, parece haber un efecto del factor "maestro" dentro de cada modalidad. También se observa un efecto de la modalidad dentro de cada maestro, más pronunciado para los maestros 1 y 2 que para el maestro 3. Esto se aprecia, nuevamente, por las pendientes de cada segmento (las primeras son decrecientes y la otra creciente) o bien graficando las medias de las modalidades para cada maestro:



Este gráfico muestra más claramente que con clases abreviadas los maestros 1 y 2 obtuvieron mejores resultados que con las no abreviadas. El maestro 3 obtuvo una pequeña diferencia a favor de CN.

Igual que sobre la variable R, sobre la variable C hay un efecto de la "combinación" de ambos factores: *maestro y modalidad de la clase*.

REFERENCIAS

Botella, J., León, O. y San Martín, R. (1993). Análisis de Datos en Psicología I. Madrid: Ediciones Pirámide.

Cano de Becerra, F. (1971). Elementos de estadística al servicio de la evaluación del rendimiento. *Revista de Psicología, 16*, 61 - 77.

Caraher, T., Carraher, D. y Schlieman, A. (1998). *En la vida diez, en la escuela cero: los contextos culturales del aprendizaje de las matemáticas*. México: Siglo XXI.

Miguel-Tobal J. J. y Cano-Vindel A. R. (1998). *Inventario de Situaciones y Respuestas de Ansiedad (ISRA)*. Manual (2da edición). Madrid: TEA.

Rosenhan, D. (1988). Acerca de estar sano en un medio enfermo. En Watzlawick, P. (comp.). *La realidad inventada ¿cómo sabemos lo que creemos saber?* (pp. 99-120). España: Gedisa.

PRÁCTICA DE INTEGRACIÓN
ESTADÍSTICA DESCRIPTIVA

EJERCICIO 1

Un fabricante de calzado deportivo está interesado en imponer su marca en el mercado a través de una efectiva campaña publicitaria dirigida a los jóvenes. Para ello le encarga a un psicólogo especializado en grupos motivacionales testear dos propuestas publicitarias de su marca de manera tal de elegir aquella con mayor impacto.

El psicólogo convoca a 30 jóvenes de 18 a 25 años de ambos sexos, residentes en Ciudad de Buenos Aires y los divide al azar en dos grupos de 15. Al grupo 1 le presenta tres publicidades, dos de las cuales pertenecen a la competencia, y la tercera la propuesta publicitaria a testear que llamaremos A. Al grupo 2 le presenta las dos mismas publicidades de la competencia y la propuesta publicitaria a testear que llamaremos B. Luego solicita a cada grupo que le asigne una posición en el ranking a las publicidades testeadas con relación a las publicidades de la competencia. Los resultados obtenidos fueron:

Ranking	Grupo 1	Grupo 2
	Propuesta A	Propuesta B
1	9	5
2	4	6
3	2	4

a) i) Indique quiénes constituyen la población de individuos y quiénes la muestra de individuos
ii) Determine quiénes constituyen las poblaciones de observaciones (¿son reales o hipotéticas?) y quiénes las muestras de observaciones.

b) Mencione la variable estadística, clasifíquela e indique el nivel de medición utilizado.

c) ¿Observa alguna tendencia en las observaciones que den cuenta de la existencia de una fuente sistemática de variación? En caso afirmativo explique dicha tendencia y mencione la fuente sistemática de variación.

d) Represente con un diagrama adecuado la comparación entre las dos propuestas.

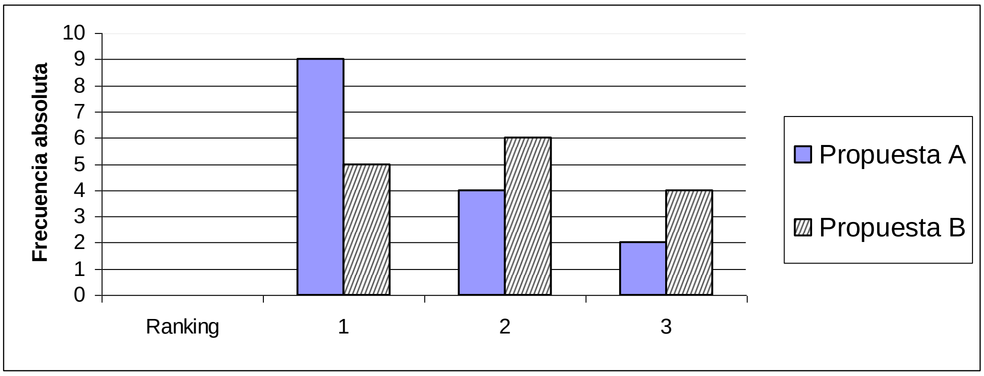
a) i. Población de individuos: Totalidad de jóvenes de 18 a 25 años de ambos sexos residentes en Ciudad de Buenos Aires. Muestra de individuos: los 30 jóvenes de 18 a 25 años de ambos sexos residentes en Ciudad de Buenos Aires seleccionados para la experiencia.

ii) Poblaciones de observaciones: una población es el conjunto de todos los rankings que los jóvenes de 18 a 25 años de ambos sexos de Ciudad de Buenos Aires le asignarían a la publicidad de haber sido expuestos a la propuesta A. La otra población es la de los rankings que dichos jóvenes le asignarían a la propuesta publicitaria B. Por cómo están definidas estas dos poblaciones, son claramente hipotéticas. Muestra de observaciones: son también dos y corresponden a cada población hipotética. Están constituidas por los rangos asignados efectivamente por los jóvenes de 18 a 25 años de ambos sexos de Ciudad de Buenos Aires seleccionados para la experiencia que vieron la publicidad tipo A y por los rangos asignados por los jóvenes que vieron la publicidad de tipo B.

b) La variable estadística es “Ranking asignado a la publicidad entre las dos marcas competidoras”. Es cuasicuantitativa y el nivel de medición utilizado es ordinal.

c) Más de la mitad de los jóvenes que recibieron la propuesta A le asignaron el primer lugar a la publicidad. El grupo de jóvenes que evaluó la propuesta B presenta opiniones más heterogéneas, no observándose ninguna posición privilegiada para esta propuesta. De modo que la tendencia es posicionar mejor la publicidad de la propuesta A. Por tanto el tipo de propuesta publicitaria parece ser una fuente sistemática de variación.

d)



EJERCICIO 2

Los siguientes datos corresponden a la cantidad de errores cometidos por un niño disléxico en la lectura de 15 párrafos breves de diferentes textos.

3	4	7	3	4	5	3	4	6	5	2	1	2	4	5

a) Presente la información en una tabla de frecuencias absolutas y calcule las medidas de tendencia central, los cuartiles, el desvío estándar.
b) ¿En qué porcentaje de ocasiones el niño cometió como máximo cuatro errores?

a)
Cálculo de la media aritmética:

X_i	n_i	**Xi·ni**
1	1	1
2	2	4
3	3	9
4	4	16
5	3	15
6	1	6
7	1	7
Total	**15**	**58**

$$\bar{x} = \frac{\sum X_i \cdot n_i}{n} \qquad \Rightarrow \qquad \bar{x} = \frac{58}{15} = 3{,}8\widehat{6}$$

Media = **3,87**

Cálculo del desvío estándar:

X_i	**n_i**	$\left(x_i - \bar{x}\right)^2 \cdot n_i$
1	1	8.2369
2	2	6.9938
3	3	2.2707
4	4	0.0676
5	3	3.8307
6	1	4.5369
7	1	9.7969
Total	**15**	**35.7335**

$$S^2 = \frac{35{,}7335}{15} = 2{,}3822 \qquad \Rightarrow \quad S = \sqrt{2{,}3822} \approx 1{,}5434$$

Cálculo de la mediana:

Para encontrar la mediana calculamos las frecuencias porcentuales acumuladas en ambos sentidos y buscamos la puntuación que supera a no más del 50% de las observaciones y es superada por no más del 50% de las mismas; la puntuación que cumple con estas condiciones es 4, pues el 40% de los niños cometió menos de 4 errores y el 33,33% de ellos cometió más de 4.

X_i	n_i	pi	P_i%	f%	g%
1	1	0.07	6.67	6.67	100.00
2	2	0.13	13.33	20.00	93.33
3	3	0.20	20.00	**40.00**	80.00
4	**4**	**0.27**	**26.67**	**66.67**	**60.00**
5	3	0.20	20.00	86.67	**33.33**
6	1	0.07	6.67	93.33	13.33
7	1	0.07	6.67	100.00	6.67

La moda es también 4 (valor de mayor frecuencia absoluta).

Cálculo de los cuartiles:

El Q_2 es 4 (coincide con la mediana).

El Q_1 es la puntuación que supera a no más del 25% de las observaciones y es superada por no más del 75% de las mismas. Dicha puntuación es 3, el 20% de los niños cometió menos de 3 errores y el 60% de ellos cometió más de 3.

El Q_3 es la puntuación que supera a no más del 75% de las observaciones y es superada por no más del 25% de las mismas. Dicha puntuación es 5, el 66,67% de los niños cometió menos de 5 errores y el 13,33% de ellos cometió más de 5.

La siguiente tabla ilustra lo que hemos explicado:

	X_i	n_i	pi	P_i%	f%	g%
	1	1	0.07	6.67	6.67	100.00
	2	2	0.13	13.33	20.00	93.33
Q_1 →	**3**	**3**	**0.20**	**20.00**	**40.00**	**80.00**
	4	4	0.27	26.67	66.67	60.00
Q_3 →	**5**	**3**	**0.20**	**20.00**	**86.67**	**33.33**
	6	1	0.07	6.67	93.33	13.33
	7	1	0.07	6.67	100.00	6.67

$\bar{x} = 3{,}87$	Mo = 4	s = 1,54
Md = 4	$Q_1 = 3$	$Q_3 = 5$

b) Para calcular el porcentaje de niños que cometió como máximo 4 errores debemos sumar el porcentaje de niños que cometió 1, 2, 3, o 4 errores, es decir:

6,67 + 13,33 + 20 + 26,67 = <u>**66,67**</u>

EJERCICIO 3

Un investigador está realizando un estudio sobre la producción académica de los profesores pertenecientes a Universidades privadas y públicas de la ciudad de Buenos Aires. Uno de los indicadores que toma para evaluar la producción es la número de participaciones en Congresos nacionales e internacionales durante el último año. Para ello toma una muestra aleatoria de 30 profesores de universidades privadas y 15 de universidades públicas y obtiene los resultados que siguen:.

Número de participaciones en Congresos	Número de Profesores	
	Universidades Privadas	Universidades Públicas
0	6	-
1	10	4
2	7	2
3	4	3
4	2	3
5	-	2
6	-	1
7	1	-

a) i) Indique quiénes constituyen la población de individuos y quiénes la muestra de individuos
ii) Determine quiénes constituyen las poblaciones de observaciones (¿son reales o hipotéticas?) y quiénes las muestras de observaciones.

b) Mencione la variable estadística, clasifíquela e indique el nivel de medición utilizado.

c) Explique qué tendencia siguen las observaciones que da cuenta de la existencia de una fuente sistemática de variación y mencione dicha fuente. ¿Todas las observaciones responden a lo previsto? ¿A qué atribuye esto? Explique.

d) Indique cuál es el gráfico adecuado para representar la distribución de frecuencias para cada tipo de universidad de modo que puedan ser comparadas.
e) ¿Qué porcentaje de profesores participó por lo menos de dos Congresos en el último año?

a)

i. Población de individuos: La totalidad de profesores de universidades privadas y públicas de la ciudad de Buenos Aires. Muestra de individuos: los 45 profesores de universidades privadas y públicas de la Ciudad de Buenos Aires.

ii. Las poblaciones de observaciones: son dos, una es la formada por todos los números que indican cuántas veces participó de un Congreso en el último año cada profesor de Universidad privada de la Ciudad de Buenos Aires. La otra población está definida de la misma manera para los profesores de universidades públicas. Son poblaciones reales. Hay dos muestras de observaciones que se corresponde con cada una de las dos poblaciones de observaciones: una es la de los 30 números que representan cuántas veces participó de un Congreso en el último año cada uno de los profesores de Universidades privadas y la otra es la definida de manera análoga sobre los 15 profesores de Universidades públicas.

b) La variable estadística es “Número de participaciones en Congresos nacionales e internacionales realizados por los profesores en el último año”. Es una variable cuantitativa discreta y el nivel de medición es de razón.

c) Para poder comparar ambos grupos es necesario usar las frecuencias porcentuales o relativas, ya que son dos muestras de diferente tamaño. Puede observarse que las frecuencias porcentuales tienden a acumularse en los valores más bajos de la variable para el caso de los profesores de Universidades privadas y hacia los valores más altos en el otro grupo. Por lo que el tipo de Universidad parece ser una fuente sistemática de variación, los profesores de las Universidades privadas tienden a participar menos que los de la Universidad pública en Congresos.
No todas las observaciones responden a esta tendencia, ya que vemos que el único profesor que más ha participado entre todos (7 presentaciones en Congresos) es de la Universidad privada. Esto puede atribuirse a fuentes fortuitas de variación.

d) Hay que realizar un diagrama de barras adyacentes usando frecuencias porcentuales.

e) Sobre el total de 45 profesores, 25 participaron por lo menos de dos Congresos en el último año. Calculamos el porcentaje que representan:

45 profesores ———— 100%
25 profesores ———— x%

$$x = \frac{25 \cdot 100}{45} \approx 55{,}56$$

El 55,56%.

EJERCICIO 4

A continuación presentamos una tabla con la distribución de frecuencias de las calificaciones finales en Lengua de niños de 7° grado del turno mañana y también algunos resúmenes estadísticos correspondientes al turno tarde.

Turno Mañana	
Calificación	n_i
8,5 – 10,5	2
6,5 – 8,5	7
4,5 – 6,5	8
2,5 – 4,5	2
0,5 – 2,5	1

Turno Tarde
$\overline{X}$ = 5,01 Md = 5,4 s = 1,86 n = 25

a) ¿Qué grupo parece haber tenido mejor rendimiento? ¿Por qué?
b) ¿En qué grupo hay mayor variabilidad relativa a la media?

a) Se trata de una variable cuantitativa. Calculamos la media aritmética y la mediana:

Turno Mañana				
Calificación	Punto Medio	n_i	n_a	$X_i \cdot n_i$
8,5 – 10,5	9.5	2	20	19
6,5 – 8,5	7.5	7	18	52.5
4,5 – 6,5	**5.5**	**8**	**11**	**44**
2,5 – 4,5	3.5	2	3	7
0,5 – 2,5	1.5	1	1	1.5
Total		20		124

Media aritmética: $\bar{x} = \frac{124}{20} = 6{,}2$

Para el cálculo de la mediana, utilizamos la fórmula correspondiente al cálculo de centiles para variables cuantitativas continuas:

El intervalo crítico es el intervalo que contiene acumuladas 10 observaciones, es decir, el intervalo 4,5 – 6,5.

$$C_{50} = 4{,}5 + \frac{2}{8} \cdot \left(\frac{50 \cdot 20}{100} - 3 \right) \quad \Rightarrow \quad C_{50} = 6{,}25$$

El turno mañana tiene mejor rendimiento ya que, resumiendo sus puntajes con las medidas de tendencia central, vemos que son mayores que en el turno tarde:

Media = 6,2 y Md = 6,25.

b) Debemos calcular el coeficiente de variación en ambos grupos. Recordemos que el coeficiente de variación nos informa sobre la representatividad de la media aritmética del conjunto de observaciones. A mayor coeficiente de variación, menor representatividad de la media, y, a menor coeficiente de variación, mayor representatividad de la misma.

Calculamos la desviación típica para los niños de ambos turnos:

Turno Mañana			
Calificación	Punto Medio	n_i	$\left(x_i - \overline{x}\right)^2 \cdot n_i$
8,5 – 10,5	9.5	2	21.78
6,5 – 8,5	7.5	7	11.83
4,5 – 6,5	5.5	8	3.92
2,5 – 4,5	3.5	2	14.58
0,5 – 2,5	1.5	1	22.09
Total		20	74.2

$$S^2 = \frac{74{,}2}{20} = 3{,}71 \quad \Rightarrow \quad S = \sqrt{3{,}71} \approx 1{,}9561$$

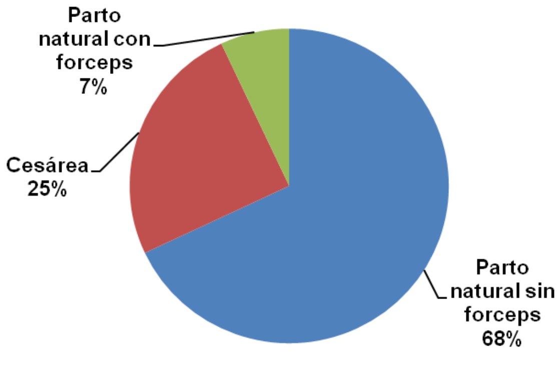
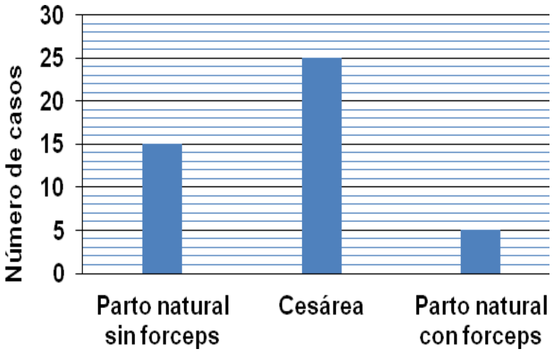
Turno Mañana: $\text{CV} = \dfrac{1{,}9561}{6{,}2} \cdot 100 \quad \Rightarrow \quad \text{CV} \approx 31{,}06$

Turno Tarde: $\text{CV} = \dfrac{1{,}86}{5{,}01} \cdot 100 \quad \Rightarrow \quad \text{CV} \approx 37{,}12$

La mayor variabilidad relativa a la media se da en el turno tarde, pues el coeficiente de variación es 37% mientras que en el turno mañana es de 31%.

EJERCICIO 5

Se realizó una investigación con el objetivo de conocer las modalidades de partos que se dan en hospitales públicos y clínicas privadas. Para ello se seleccionó una muestra aleatoria de 100 madres que dieron a luz en hospitales y clínicas emplazadas en jurisdicción del Gobierno de la Ciudad de Buenos Aires durante el mes de setiembre de 2006. A continuación se presenta la información obtenida:

Hospitales Públicos	**Clínicas Privadas**
	

a) Determine la población y la muestra de individuos.
b) Mencione la variable de interés y clasifíquela. Mencione qué nivel de medición se utilizó e indique los valores de la escala.
c) Mencione la posible fuente sistemática de variación.
d) ¿Qué tipo de gráficos se han utilizado para representar la información de ambas distribuciones?

a) Población de Individuos: todas las madres que han dado a luz en hospitales públicos y clínicas privadas emplazadas en jurisdicción del Gobierno de la Ciudad de Buenos Aires, durante el mes de setiembre de 2006. Muestra de Individuos: 100 madres que han dado a luz en hospitales públicos y clínicas privadas emplazadas en jurisdicción del Gobierno de la Ciudad de Buenos Aires durante el mes de setiembre de 2006.

b) La variable es “Modalidad de parto”, es de tipo cualitativa. Se ha empleado una escala nominal cuyos valores son: “Parto Natural”, “Cesárea” y “Fórceps”.

c) Una posible fuente sistemática de variación es: Tipo de establecimiento donde se realizó el parto: hospitales públicos y clínicas privadas.

d) Para representar la distribución del Hospital Público se ha utilizado un pictograma en tanto para la de la Clínica Privada un diagrama de rectángulos.

EJERCICIO 6

Con la misma información que se presenta en el Ejercicio 5:

a) Determine qué cantidad de mujeres corresponde a la muestra de hospitales públicos y clínicas privadas, respectivamente, observando ambos gráficos.

b) Presente en dos tablas cada una de las distribuciones de frecuencias, especificando en cada caso las absolutas, relativas y porcentuales.

c) ¿Con qué medidas compararía las dos distribuciones con respecto a su variabilidad y a su tendencia central? Aplíquelas y extraiga conclusiones.

a) En las clínicas privadas dieron a luz un total de 45 mujeres: 15 por parto natural, 25 con cesárea y 5 con fórceps.

Siendo la muestra seleccionada de 100 mujeres, en hospitales públicos dieron a luz un total de 55 mujeres.

b) ***Hospitales Públicos***:

La información presente en el gráfico para la modalidad de parto en Hospitales Públicos, corresponde a las frecuencias porcentuales. Dividiendo cada una de ellas por 100, se obtienen las frecuencias relativas para cada categoría, es decir, 0,68; 0,25 y 0,07, respectivamente.

Para obtener las frecuencias absolutas multiplicamos la frecuencia relativa por ***n***:

Parto natural, 68%: 0,68 · 55 = 37,4 (37 mujeres)

Cesárea, 25%: 0,25 · 55 = 13,75 (14 mujeres)

Fórceps, 7%: 0,07 · 55 = 3,85 (4 mujeres)

Clínicas Privadas:

Las frecuencias relativas en ambas modalidades se obtienen dividiendo las frecuencias absolutas de cada categoría por el tamaño de la muestra.

Parto natural: $\frac{15}{45} = 0{,}33$

Cesárea: $\frac{25}{45} = 0{,}56$

Fórceps: $\dfrac{5}{45} = 0{,}11$

Las frecuencias porcentuales se obtienen multiplicando las frecuencias relativas por 100. Las mismas son 33.3, 55.6 y 11.1,, respectivamente.

Resumimos los cálculos realizados en la siguiente tabla:

	Hospitales Públicos			Clínicas Privadas		
X_i	n_i	p_i	P_i%	n_i	p_i	P_i%
Parto Natural	37	0.68	68	15	0,33	33,3
Cesárea	14	0,25	25	25	0,56	55,6
Fórceps	4	0,07	7	5	0,11	11,1
	55	1	100	45	1	100

c) Como se trata de una variable medida en escala nominal la única medida de tendencia central que es posible emplear es la Moda. Para la distribución de los Hospitales Públicos la Moda es “Parto Natural” y para la distribución de las Clínicas Privadas es “Cesárea”. La variabilidad de las distribuciones puede ser comparada con la Entropía. Como es posible observar, la moda de la distribución de los Hospitales Públicos concentra un mayor número de casos (68%) que la moda de las Clínicas Privadas (55,6%), esto nos indica que esta última es la distribución con mayor entropía durante el mes de setiembre de 2006.

EJERCICIO 7

Un grupo de investigadores estudia la incidencia de la cafeína sobre la memoria auditiva. Para ello reunieron dos grupos, una denominado “experimental” conformado por 15 personas, a las que se les suministró 600 miligramos de cafeína. Otro, denominado “de control”, conformado por 26 personas, a las que no se les suministró sustancia alguna. Ambos grupos fueron sometidos a una prueba de memoria auditiva, los resultados se presentan a continuación:

Grupo Experimental	Cantidad de palabras olvidadas	n_i
	10	2
	9	6
	8	5
	7	0
	6	2

Grupo Control	$\overline{X} = 6$ Md = 6,5 Mo = 6 $s^2 = 1{,}4$ n = 26

Mediante el cálculo de los estadísticos adecuados responda y justifique:

a) ¿Parece influir el consumo de cafeína en la memoria auditiva a juzgar por los resultados muestrales?

b) ¿Qué grupo tiene menor variabilidad relativa?

c) Calcule el tercer cuartil de la distribución del grupo de experimental.

d) Los investigadores desean seleccionar a aquellos individuos del grupo experimental que no hayan superado el primer cuartil. ¿Cuántas personas cumplen esta condición?

a) Calculamos la media para el grupo experimental:

Cantidad de palabras olvidadas	n_i	$n_i \cdot x_i$
10	2	20
9	6	54
8	5	40
7	0	0
6	2	12
Total	**15**	**126**

$$\bar{x} = \frac{126}{15} = 8,4$$

Al comparar las medias obtenidas por ambos grupos se encuentra que el Grupo Experimental (media = 8,4) ha tenido un peor rendimiento que el Grupo Control (media = 6). Si bien no nos es posible determinar aún si esta diferencia es significativa, los datos muestrales apuntan en la dirección de la hipótesis de que el consumo de cafeína afecta negativamente el rendimiento de la memoria.

b) Para determinar qué grupo tiene menor variabilidad relativa, debemos calcular el coeficiente de variación en ambos grupos.

Grupo Experimental:

Cantidad de palabras olvidadas	n_i	$(x_i - \bar{x})^2 \cdot n_i$
10	2	5,12
9	6	2,16
8	5	0,8
7	0	0
6	2	11,52
Total	**15**	**19,6**

Varianza: $S^2 = \frac{19,6}{15} = 1,3067$

Desviación Típica: $S = \sqrt{1,3067} \approx 1,1431$

Coeficiente de Variación: $CV = \frac{1,1431}{8,4} \cdot 100 \approx 13,61$

Grupo de Control:

Desviación Típica: $S = \sqrt{1,4} \approx 1,1832$

Coeficiente de Variación: $CV = \frac{1,1832}{6} \cdot 100 \approx 19,72$

El Grupo Experimental presenta menor variabilidad relativa ya que su CV es 13,6%, en tanto que el Grupo Control tiene un CV de 19,7%.

c) Calcular el Q_3 es equivalente a calcular el C_{75}; es la puntuación que *supera a no más del 75%* de las observaciones y que es *superada por no más del 25%* de las mismas.

Calculamos las frecuencias porcentuales acumuladas en ambos sentidos:

Cantidad de palabras olvidadas	n_i	P_i%	f_a%	g_a%
10	2	13,33	100	**13,33**
9	**6**	**40,00**	**86,67**	**53,33**
8	5	33,33	**46,67**	86,67
7	0	0,00	13,33	86,67
6	2	13,33	13,33	100

El valor que cumple la condición de superar a no más del 75% de las observaciones y ser superado por no más del 25% de las mismas es 9 (el 46,67% de los sujetos olvidó menos de 9 palabras y el 13,33% olvidó más de 9 palabras).

Q_3 = 9 palabras olvidadas.

d) El Q_1 es la puntuación que *supera a no más del 25%* de las observaciones y que es *superada por no más del 75%* de las mismas.

Cantidad de palabras olvidadas	n_i	P_i%	f_a%	g_a%
10	2	13,33	100	13,33
9	6	40,00	86,67	**53,33**
8	**5**	**33,33**	**46,67**	**86,67**
7	0	0,00	**13,33**	86,67
6	2	13,33	13,33	100

El valor que cumple la condición de superar a no más del 25% de las observaciones y ser superado por no más del 75% de las mismas es 8 (el 13,33% de los sujetos olvidó menos de 8 palabras y el 53,33% olvidó más de 8 palabras).

Q_1 = 8 y, en consecuencia, serán 7 las personas seleccionadas (2 + 5 personas no superan las 8 palabras olvidadas).

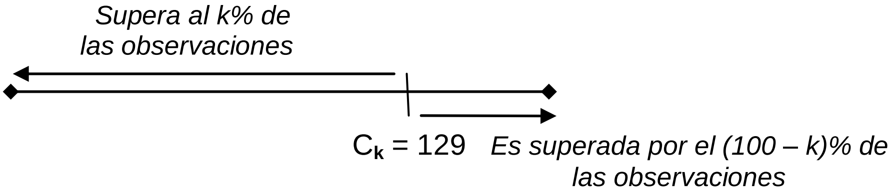
EJERCICIO 8

Un establecimiento educativo utiliza el cociente intelectual de Binet – Stern para ubicar a los alumnos aspirantes al ingreso en tres grupos. El grupo 1 estará integrado por aquéllos que no superen el centil 15 y el grupo 3 tiene como mínimo puntaje un CI de 129. A continuación se presenta la distribución de frecuencias correspondiente:

CI	n_i
136 - 150	8
121 - 135	24
106 - 120	36
91 - 105	11
76 - 90	1

a) Calcule el porcentaje de alumnos que ingresarán al grupo 3.
b) ¿Cuál será la puntuación mínima para ingresar al grupo 2?
c) Si los directivos desean seleccionar a quienes superen el centil 95 o no superan el centil 5 ¿Cuántos alumnos serán elegidos?

a) El grupo 3 está formado por aquellos alumnos que tienen como mínimo puntaje un CI de 129. Calculamos, en primer lugar, el porcentaje de observaciones que deja por debajo de sí la puntuación 129. Para ello utilizamos la fórmula para el cálculo de centiles, cuya incógnita es *k*.



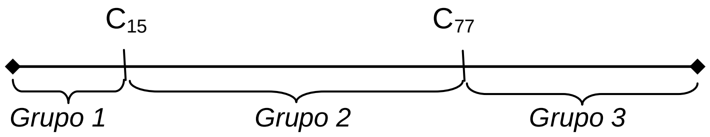
CI	n_i	Frecuencia absoluta acumulada
136 - 150	8	80
121 - 135	**24**	**72**
106 - 120	36	48
91 - 105	11	12
76 - 90	1	1

El intervalo crítico es aquél que contiene a la puntuación 129. Dicho intervalo es 121 – 135.

$$129 = 120{,}5 + \frac{15}{24}\cdot\left(\frac{k\cdot 80}{100} - 48\right) \quad \Rightarrow \quad k = \left[\left(129 - 120{,}5\right) : \frac{15}{24} + 48\right]\cdot\frac{100}{80} = 77$$

El porcentaje de alumnos que supera la puntuación 129 es 77%, por lo tanto ingresarán al grupo 3 el 23% de ellos (100 – 77 = 23).

b) Los grupos están conformados de la siguiente forma:



El grupo 2 estará formado por aquellos alumnos que tengan un CI comprendido entre el C_{15} y el C_{77}, por lo tanto, la puntuación mínima que debe tener un alumno para ingresar al mismo es la que corresponde al C_{15}.

El 15% de 80 es: 0,15 · 80 = 12

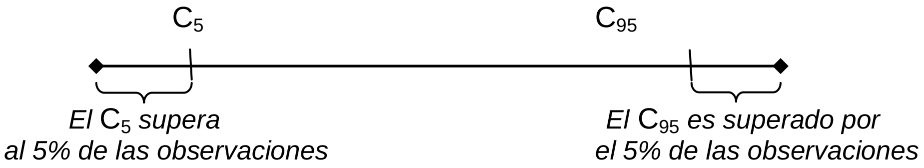
El intervalo crítico es el que contiene acumuladas 12 observaciones, es decir, el intervalo 91 – 105:

CI	n_i	Frecuencia absoluta acumulada
136 - 150	8	80
121 - 135	24	72
106 - 120	36	48
91 - 105	**11**	**12**
76 - 90	1	1

Observamos que, en este caso, este intervalo (90,5 – 105,5) contiene exactamente 12 observaciones acumuladas que coincide con el equivalente al porcentaje que estamos buscando. Esto significa que la puntuación 105,5 supera exactamente a 12 observaciones, con lo cual no es necesario aplicar la fórmula de centiles para el cálculo del C_{15}. La puntuación correspondiente al C_{15} es el límite exacto superior del intervalo crítico, es decir, 105,5.

El puntaje mínimo para ingresar al grupo 2 es un CI de 105,5.

c) Se desea seleccionar a quienes superen el centil 95 o no superan el centil 5.



El centil 95 es la puntuación superada por el 5% de las observaciones, y el centil 5 es la puntuación que supera al 5% de las mismas. Por lo tanto, se selecciona en total al 10% de los alumnos.

El 10% de 80 es: 0,10 · 80 = 8

Deben seleccionar a 8 niños.

EJERCICIO 9

Las autoridades de nuestra facultad están interesadas en conocer el tiempo que tardan los alumnos en llegar al establecimiento ya sea desde sus hogares o lugares de trabajo. Se solicita al alumno que considere el lugar más frecuente desde el cual se traslada para llegar a la facultad y que mida el tiempo requerido. Para tal fin, fueron seleccionados al azar 20 alumnos de los tres turnos que concurren a la sede de HY y 20 alumnos de los tres turnos que concurren a la sede de Independencia. A continuación se presenta la distribución de frecuencias de una sede y algunos estadísticos correspondientes a la otra:

Sede Hipólito Yrigoyen	Tiempo (minutos)	n_i
	60 – 74	1
	45 – 59	6
	30 – 44	3
	15 – 29	8
	0 – 14	2

Sede Independencia	$\bar{X} = 35$ $s^2 = 81$ $n = 20$

a) Mencione la población y muestra de individuos.
b) Mencione la población y muestra de observaciones.
c) Mencione la variable de interés y clasifíquela. Indique qué nivel de medición se utilizó.

a) <u>Población de Individuos</u>: dos poblaciones de individuos, todos los estudiantes que asisten a la sede de HY y todos los estudiantes que asisten a la sede de Independencia de la Facultad de Psicología de la UBA. <u>Muestra de individuos</u>: dos muestras de individuos, 20 estudiantes que asisten a la sede de HY y 20 estudiantes que asisten a la sede de Independencia de la Facultad de Psicología de la UBA.

b) <u>Población de observaciones</u>: hay dos, el conjunto de números que representan al tiempo expresado en minutos que tardan en llegar (desde sus casas o lugares de trabajo) los estudiantes que asisten a la sede de HY y el conjunto de números que representan al tiempo expresado en minutos que tardan en llegar (desde sus casas o lugares de trabajo) los estudiantes que concurren a la sede de Independencia, de la Facultad de Psicología, UBA.
<u>Muestra de observaciones</u>: hay dos, el conjunto de números que representan al tiempo expresado en minutos que tardan en llegar (desde sus casas o lugares de trabajo) los 20 alumnos que concurren a la sede de HY y el conjunto de números que representan al tiempo expresado en minutos que tardan en llegar (desde sus casas o lugares de trabajo) los 20 alumnos que concurren a la sede de Independencia, Facultad de Psicología, UBA.

c) Tiempo que tarda un estudiante en llegar a la Facultad de Psicología (sede HY o Independiencia) desde su hogar o lugar de trabajo. Variable cuantitativa continua. Nivel de medición de cociente o razón.

EJERCICIO 10

A partir de la información suministrada en el Ejercicio 9 determine:
a) Para qué sede el tiempo que tardan los alumnos en llegar es mayor. Justifique.
b) Cuál de las dos medias es más representativa. Justifique la respuesta.

b) Calculamos el tiempo medio que tardan los alumnos en llegar al establecimiento para la sede de Hipólito Yrigoyen:

Tiempo (minutos)	x_i	n_i	$x_i \cdot n_i$
60 – 74	67	1	67
45 – 59	52	6	312
30 – 44	37	3	111
15 – 29	22	8	176
0 – 14	7	2	14
Total		**20**	**680**

$$\bar{x} = \frac{680}{20} = 34$$

El tiempo medio que tardan en llegar los alumnos al establecimiento, en HY, es 34 minutos. Los alumnos tardan más en llegar a la sede de Independencia pues el tiempo medio correspondiente a esta sede es mayor, 35 minutos.

b) Para determinar cuál de las dos medias es más representativa, debemos hacerlo en términos de la variabilidad de cada grupo, como ya dijimos anteriormente, a mayor coeficiente de variación, menor representatividad de la media y, recíprocamente, menor coeficiente de variación, mayor representativa es la media.

Calculamos el coeficiente de variación en ambos grupos.

<u>Sede Hipólito Yrigoyen</u>:

Tiempo (minutos)	x_i	n_i	$(x_i - \bar{x})^2 \cdot n_i$
60 – 74	67	1	1089
45 – 59	52	6	1944
30 – 44	37	3	27
15 – 29	22	8	1152
0 – 14	7	2	1458
Total		**20**	**5670**

$$S^2 = \frac{5670}{20} = 283{,}5 \qquad \Rightarrow \quad S = \sqrt{283{,}5} \approx 16{,}837$$

$$CV = \frac{16{,}837}{34} \cdot 100 \approx 49{,}52$$

En HY el CV = 49,52%.

<u>Sede Independencia</u>:

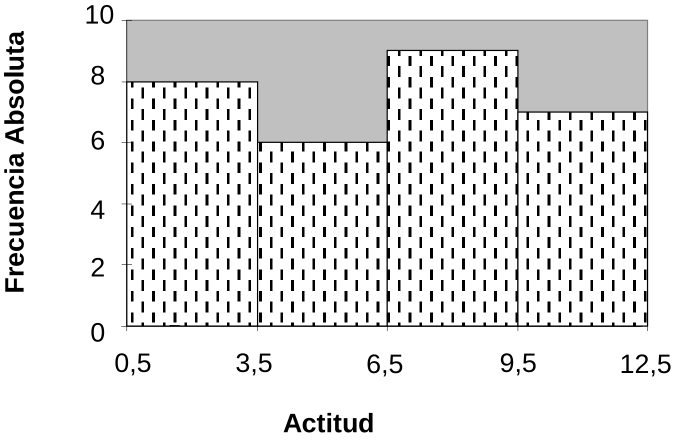
En Independencia: S = 9 y la media es 35.

$$CV = \frac{9}{35} \cdot 100 \approx 25{,}7$$

El CV = 25.7%, menor que en la sede HY. La media correspondiente a la sede de Independencia es más representativa porque existe menor variabilidad relativa.

EJERCICIO 11

Un grupo de 30 mujeres jubiladas elegidas aleatoriamente de los Centros de Jubilados del barrio de Flores respondió a una escala de actitud hacia el consumo de medicamentos no prescriptos por un médico. Esta escala permite discriminar a los sujetos propensos a consumir de manera abusiva si superan los 11 puntos. El gráfico resume los resultados obtenidos.



a) Mencione la población y la muestra de individuos.
b) Mencione la población y muestra de observaciones.
c) Mencione la variable, clasifíquela conforme al tratamiento estadístico que se le ha dado e indique el nivel de medición.

a) Población individuos: las mujeres jubiladas que concurren a los Centros de Jubilados del barrio de Flores. Muestra de individuos: 30 mujeres jubiladas que concurren a los Centros de Jubilados del barrio de Flores.

b) Población de observaciones: el conjunto de puntajes que podrían obtener en esta escala todas las mujeres jubiladas que concurren a los Centros de Jubilados del barrio de Flores. Muestra de observaciones: el conjunto de puntajes obtenidos en esta escala por las 30 mujeres jubiladas que concurren a los Centros de Jubilados del barrio de Flores.

c) La variable es: “Puntaje en la actitud hacia el consumo de medicamentos no prescriptos por un médico”. Se le ha dado un tratamiento de variable continua medida en un nivel de medición utilizado es al menos intervalar; al no explicarse cómo se ha contruido la escala para saber si el origen es convencional o no, no se sabe si puede ser de razón.

EJERCICIO 12

A partir de la información del Ejercicio 11,

a) Construya la tabla de distribución de frecuencias que incluya: frecuencias absolutas, relativas, porcentuales y acumuladas (absolutas, relativas y porcentuales).
b) Indique qué otro u otros gráficos podría utilizar para representar la distribución de frecuencias de esta variable.

a) Leemos del gráfico, la frecuencia absoluta correspondiente a cada intervalo: 8 para el intervalo 0,5 – 3,5; 6 para el intervalo 3,5 – 6,5; 9 para el intervalo 6,5 – 9,5; 7 para el intervalo 9,5 – 12,5. Sus respectivos límites aparentes son: 1 – 3; 4 – 6; 7 – 8; 9 – 12.

A partir de las frecuencias absolutas calculamos las frecuencias relativas, porcentuales y sus respectivas acumuladas:

Frecuencias Relativa y Porcentual:

Para el cálculo de las frecuencias relativas, dividimos cada frecuencia absoluta por 30; las frecuencias porcentuales se obtienen multiplicando por 100 cada uno de estos valores:

$$\frac{7}{30} \approx 0{,}23\,; \quad 0{,}23 \cdot 100 = 23$$

$$\frac{9}{30} = 0{,}3\,; \qquad 0{,}3 \cdot 100 = 30$$

$$\frac{6}{30} = 0{,}2\,; \qquad 0{,}2 \cdot 100 = 20$$

$$\frac{8}{30} \approx 0{,}27\,; \qquad 0{,}27 \cdot 100 = 27$$

X_i	N_i	p_i	P_i %
10-Dic	7	0.23	23
07-Sep	9	0.3	30
04-Jun	6	0.2	20
01-Mar	8	0.27	27
Total	30	1	100

Para obtener las respectivas frecuencias acumuladas procedemos a sumar, en cada intervalo, la respectiva frecuencia más la acumulada del intervalo anterior.

A continuación se muestra la tabla completa de distribución de frecuencias:

X_i	N_i	n_a	p_i	p_a	P_i %	P_a %
10 - 12	7	30	0.23	1	23	100
7 - 9	9	23	0.3	0.77	30	77
4 - 6	6	14	0.2	0.47	20	47
1 - 3	8	8	0.27	0.27	27	27
Total	30	-	1	-	100	-

b) Polígono de frecuencias y Ojiva de Galton (polígono de frecuencias acumuladas)

EJERCICIO 13

A partir de la información de los Ejercicios 11 y 12 responda:

a) ¿Cuántas mujeres de la muestra resultaron ser propensas a consumir medicamentos de manera abusiva?, teniendo en cuenta que esta escala permite discriminar a los sujetos propensos a consumir de manera abusiva si superen los 11 puntos.

b) Si una muestra de 30 mujeres que concurren a Centros de Jubilados del barrio de Balvanera presentan una media en la escala de actitud hacia el consumo de medicamentos no prescriptos por un médico de 4,3, indique cuál de las dos muestras presenta una actitud más proclive al consumo de medicamentos no prescriptos. Justifique la respuesta.

a) Los sujetos que se consideran propensos a consumir medicamentos de manera abusiva son aquéllos que superan los 11 puntos.

Calculamos, en primer lugar, el rango percentilar correspondiente a 11 puntos. Dicha puntuación pertenece al intervalo 10 – 12 (intervalo crítico).

$$11 = 9{,}5 + \frac{3}{7} \cdot \left(\frac{k \cdot 30}{100} - 23 \right) \quad \Rightarrow \quad k = \left[\,(11 - 9{,}5) : \frac{3}{7} + 23 \,\right] \cdot \frac{100}{30} = 88{,}\widehat{3}$$

La puntuación 11 supera al 88% de las observaciones, y es superada por el 12% de las mismas. Luego, el porcentaje de sujetos que se consideran propensos a consumir medicamentos de manera abusiva es 12 (100% - 88%). El 12% de 30 mujeres es igual a 3.6 ~ 4. Hay 4 mujeres de la muestra propensas a consumir medicamentos no prescriptos por un médico de manera abusiva.

b) Para determinar cuál de las dos muestras presenta una actitud más proclive al consumo de medicamentos no prescriptos, debemos conocer la media de las 30 mujeres jubiladas de los Centros de Jubilados del barrio de Flores.

X_i	Punto Medio	Ni	$X_i \cdot N_i$
10 - 12	11	7	77
7 - 9	8	9	72
4 - 6	5	6	30
1 - 3	2	8	16
Total		**30**	**195**

$$\bar{x} = \frac{195}{30} = 6,5$$

El promedio de puntaje en la escala entre las mujeres jubiladas del barrio de Balvanera es 4.3, mientras que entre las jubiladas del barrio de Flores es de 6.5. Estas últimas presentan una actitud más favorable hacia el consumo de medicamentos no prescriptos.

EJERCICIO 14

Un investigador ha relevado la cantidad de horas que miran televisión los fines de semana los niños residentes en Ciudad de Buenos Aires, para lo cual tomó una muestra aleatoria de 16 niños. Otro investigador realizó el mismo relevamiento pero sobre una muestra de niños del Gran Buenos Aires. A continuación se presentan los datos de ambos investigadores:

Ciudad de Buenos Aires	Cantidad de horas	n_i
	12 – 14	2
	9 – 11	4
	6 – 8	5
	3 – 5	3
	0 – 2	2

Gran Buenos Aires	$\bar{X}$ = 6,3hs. s = 2,2hs. Mo = 6,8hs Md = 6,2hs n = 20

a) ¿Qué grupo mira en general mayor cantidad de horas de televisión? ¿Por qué?
b) Si se comparan las dos muestras ¿cuál presenta mayor variabilidad relativa a la media?

a) Calculamos la cantidad media de horas que miran televisión los fines de semana los niños residentes en Ciudad de Buenos Aires:

Cantidad de horas		n_i	$x_i \cdot n_i$
12 – 14	13	2	26
9 – 11	10	4	40
6 – 8	7	5	35
3 – 5	4	3	12
0 – 2	1	2	2
Total		**16**	**115**

$$\bar{x} = \frac{115}{16} = 7{,}1875$$

$\bar{x}$ = 7,19 (Ciudad de Buenos Aires). En la muestra de Ciudad de Buenos Aires se mira mayor cantidad de horas televisión dado que la media en esta localidad es mayor que en Gran Buenos Aires (6,3).

b) En Ciudad de Buenos Aires el s = 3.59 y el CV = 49.98%. En Gran Buenos Aires: s = 2.2 y el CV = 34.92%. En la distribución correspondiente a la Ciudad de Buenos Aires existe mayor variabilidad relativa a la media.

PRÁCTICA IV

La comparación entre puntuaciones directas puede llevar a conclusiones engañosas, es por eso que en la práctica se suele transformar las puntuaciones observadas en otras que, sin perder o distorsionar la información contenida en las puntuaciones originales, permitan una comparación directa de las mismas. Los instrumentos desarrollados para ello son:

- ***Puntuaciones Típicas***: La puntuación típica de una observación indica el número de desviaciones típicas que esa observación se separa de la media del grupo de observaciones. Su media es 0 y su varianza y su desviación típica son iguales a 1. Su fórmula es:

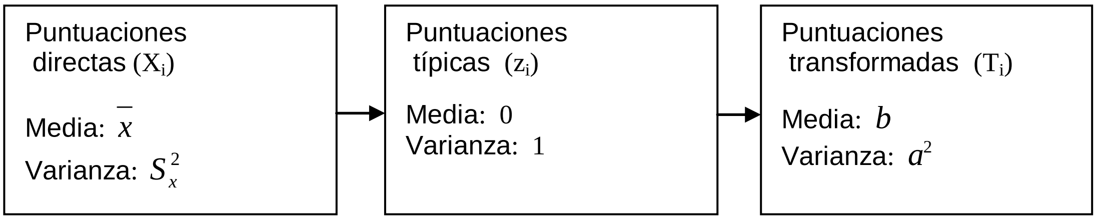
$$z_i = \frac{X_i - \overline{x}}{S_x}$$

- ***Escalas derivadas***: Dado que buena parte de las puntuaciones típicas suelen ser negativas y casi todas decimales, muchas veces se transforman las mismas, mediante una función lineal, en otras puntuaciones equivalentes a ellas, pero en las que se evita la dificultad operativa. Estas puntuaciones constituyen lo que se denomina una escala derivada. La fórmula de transformación es:

$$T_i = a \cdot z_i + b$$

con $a \neq 0$, donde $\overline{T} = b$, $S_T = |\, a \,|$ y $S_T^2 = a^2$.

Esquema de transformación en una escala derivada (Botella, página 159 – Cuadro 6.1)



Tipificación

$$z_i = \frac{X_i - \overline{x}}{S_x}$$

Transformación en escala derivada

$$T_i = a \cdot z_i + b$$

Aunque las constantes a y b pueden tomar cualquier valor, se han propuesto entre otras las siguientes escalas con nombre propio:

$$T = 10z + 50$$

$$S = 2z + 5$$

$$CI = 15z + 100$$

EJERCICIO 1

Complete la siguiente tabla a partir de la información exhibida.

	TEST A			TEST B			TEST C			
	Puntaje	Media	Desv. Est.	Puntaje	Media	Desv. Est.	Puntajes		Media	Desv. Est.
Bruto	65	45	8	40		12	36	18		
Z				-1				-2		
T							60			
CI										

TEST A

$$z_i = \frac{X_i - \bar{x}}{S_x} \quad \Rightarrow \quad z = \frac{65-45}{8} = 2{,}5$$

La media de las puntuaciones típicas es 0 y la desviación típica es 1.

$$T = 10z + 50 \quad \Rightarrow \quad T = 10 \cdot 2{,}5 + 50 = 75$$

La media de las puntuaciones T es 50 y la desviación típica es 10.

$$CI = 15z + 100 \quad \Rightarrow \quad CI = 15 \cdot 2{,}5 + 100 = 137{,}5$$

La media de las puntuaciones CI es 100 y la desviación típica es 15.

TEST B

El puntaje $X_i = 40$ se encuentra una desviación típica por debajo de la media (el puntaje z que le corresponde es -1), con lo cual $\bar{x} = 40 + 12 = 52$

$$T = 10z + 50 \quad \Rightarrow \quad T = 10 \cdot (-1) + 50 = 40$$

$$CI = 15z + 100 \quad \Rightarrow \quad CI = 15 \cdot (-1) + 100 = 85$$

TEST C

El puntaje $X_i = 18$ se encuentra dos desviaciones típicas por debajo de la media (el puntaje z que le corresponde es -2), con lo cual:

$$\bar{x} - 2S = 18 \qquad (1)$$

Según vimos en el esquema de transformación de una escala derivada, las puntuaciones directas se tipifican y luego se transforman en puntajes T. Haciendo el proceso inverso, a partir de un puntaje T, podemos obtener la puntuación típica correspondiente y luego proceder como en el caso anterior.

El puntaje T que le corresponde a $X_i = 36$ es 60, con lo cual:

$$T = 10z + 50 \quad \Rightarrow \quad 60 = 10 \cdot z + 50 \quad \Rightarrow \quad z = \frac{60-50}{10} = 1$$

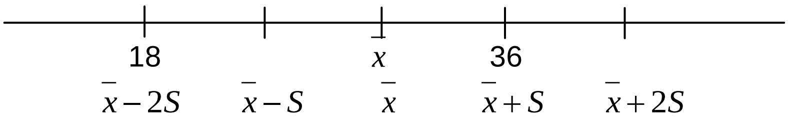
La puntuación típica correspondiente a $X_i = 36$ es 1, es decir, dicha puntuación se encuentra una desviación típica por encima de la media.

Por lo que,

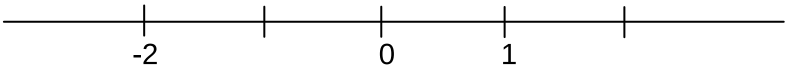
$$\bar{x} + S = 36 \qquad (2)$$

Un esquema que representa las dos situaciones anteriores es el siguiente:

Puntaje X:



Puntaje z:



Resolviendo el sistema formado con las ecuaciones (1) y (2), encontramos los valores de $\bar{x}$ y S:

$$\begin{cases} \bar{x} - 2S = 18 \\ \bar{x} + S = 36 \end{cases} \Rightarrow \begin{cases} \bar{x} = 18 + 2S \\ \bar{x} = 36 - S \end{cases} \Rightarrow 18 + 2S = 36 - S \Rightarrow 2S + S = 36 - 18$$

$$3S = 18 \Rightarrow S = 6$$

$$\bar{x} = 36 - 6 \Rightarrow \bar{x} = 30$$

Completamos los valores restantes:

$$T = 10z + 50 \Rightarrow T = 10 \cdot (-2) + 50 = 30$$

$$CI = 15z + 100 \Rightarrow CI = 15 \cdot 1 + 100 = 115$$

$$CI = 15 \cdot (-2) + 100 = 70$$

Resumimos los cálculos realizados en la siguiente tabla:

	TEST A			TEST B			TEST C			
	Puntaje	Media	Desv. Est.	Puntaje	Media	Desv. Est.	Puntajes		Media	Desv. Est.
Bruto	**65**	**45**	**8**	**40**	52	**12**	**36**	**18**	30	6
Z	2,5	0	1	**-1**	0	1	1	**-2**	0	1
T	75	50	10	40	50	10	**60**	30	50	10
CI	137,5	100	15	85	100	15	115	70	100	15

EJERCICIO 2

Indique cuáles de las siguientes afirmaciones son verdaderas y cuáles no.

a) En una prueba deportiva de velocidad se registra la distancia en metros recorrida por ciertos atletas en un minuto. En esta prueba Andrés obtuvo un puntaje Z = - 0,2. Posteriormente se exhibió la distancia recorrida por cada atleta en centímetros. Entonces, en esta escala, el puntaje Z de Andrés resultó ser - 20.

b) Los puntajes T siempre se hallan comprendidos entre 20 y 80.
c) La valoración de las puntuaciones tipificadas, en principio, sólo es posible en términos relativos al grupo de referencia y no en términos absolutos.
d) Los puntajes Z siempre se hallan comprendidos entre -1 y 1.
e) Al medir en segundos el tiempo de reacción a un estímulo, Catalina obtiene un puntaje Z = 0,5. Eso significa que tardó en reaccionar medio segundo más que la media del grupo normativo.
f) Las puntuaciones tipificadas son adimensionales y por eso permanecen invariantes si se cambian las unidades de la escala.

a) **Falsa**. Las puntuaciones típicas son adimensionales.

b) **Falsa**. Es poco probable encontrar puntajes T menores a 20 o mayores a 80, ya que el 99,74% de la población se encuentra entre $\mu \pm 3\sigma$, o, en puntuaciones típicas, esto es entre – 3 y 3. Pero esto no significa que en la escala los valores de T estén comprendidos entre 20 y 80. Para valores de z menores a – 3 o mayores a 3, encontramos valores T menores a 20 o mayores a 80, respectivamente.

c) **Verdadera**. Se trata de puntuaciones relativas al grupo de referencia.

d) **Falso**.

e) **Falso**. Las puntuaciones típicas indican cuántas desviaciones típicas se separa una puntuación de la media del grupo; en este caso, z = 0,5 indica sólo que la posición de Catalina respecto de la media del grupo es media desviación típica por encima de la misma.

f) **Verdadera**. Al traducir todas las puntuaciones directas a puntuaciones típicas llevamos los resultados a una escala común, cualquiera que sea la magnitud de las puntuaciones originales.

EJERCICIO 3

Analía tiene un puntaje Z = 2 en Perfeccionismo y T = 35 en Afabilidad, dos escalas del Cuestionario Factorial de Personalidad 16PF. Los puntajes Z y T de Gonzalo en dichas escalas son respectivamente -1,5 y 20. En estas pruebas, puntajes más altos representan mayores niveles del rasgo medido. Indique cuáles de todas las siguientes afirmaciones son verdaderas y cuáles falsas.

a) El puntaje directo de Analía en Perfeccionismo está dos puntos por encima de la media del grupo normativo.
b) Gonzalo es menos perfeccionista y afable que la media del grupo normativo.
c) El puntaje directo de Gonzalo es más alto en Afabilidad que en Perfeccionismo.
d) Analía es tan afable como Gonzalo perfeccionista.
e) No es posible comparar dos rasgos tan diferentes como son el perfeccionismo y la afabilidad.
f) Gonzalo es más perfeccionista que afable.
g) El puntaje bruto de Gonzalo en Perfeccionismo se halla a una desvío estándar y medio por debajo de la media del grupo normativo.

Para responder a alguno de los puntos anteriores, debemos transformar a una escala común los puntajes de Analía y Gonzalo en ambas características:

Analía:

Afabilidad: T = 35 $\Rightarrow$ $35 = 10z + 50$ $\Rightarrow$ $z = \frac{35-50}{10}$ $\Rightarrow$ $z = -1{,}5$

Gonzalo:

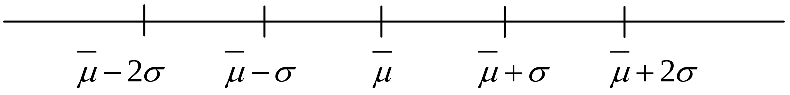
Afabilidad: T = 20 $\Rightarrow$ $20 = 10z + 50$ $\Rightarrow$ $z = \frac{20-50}{10}$ $\Rightarrow$ $z = -3$

Resumimos los datos en la siguiente tabla:

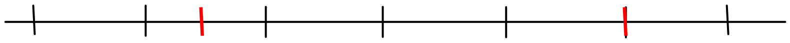
	Afabilidad	Perfeccionismo
Analía	-1,5	2
Gonzalo	-3	-1,5

Gráficamente:

Puntaje X:



Analía

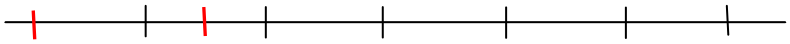


Puntaje Z: - 3 - 2 - 1 0 1 2 3

Afabilidad **- 1,5**

Perfeccionismo **2**

Gonzalo



Puntaje Z: - 3 - 2 - 1 0 1 2 3

Afabilidad **- 3**

Perfeccionismo **- 1,5**

a) ***Falsa***. El puntaje directo de Analía en Perfeccionismo está dos *desviaciones típicas* por encima de la media del grupo.

b) ***Verdadera***. Los puntajes Z de Gonzalo, en ambas características, son negativos, se encuentran por debajo de la media del grupo. Por lo tanto, es menos perfeccionista y afable que la media del grupo normativo.

c) ***Falsa***. Su puntaje Z en Afabilidad (- 1,5) es mayor que en Perfeccionismo (- 3)

d) ***Verdadera***. Ambos obtuvieron el mismo puntaje Z (- 1,5) en Afabilidad.

e) ***Falsa***. Las puntuaciones típicas, precisamente, permiten hacer comparaciones entre unidades de distintos grupos, entre variables medidas de distintas formas o, incluso, entre variables diferentes. Análogamente si hubiéramos trabajado con los puntajes T de Analía y Gonzalo, ya que son puntuaciones equivalentes.

f) ***Verdadera***. Su puntaje Z en Perfeccionismo es mayor que su puntaje Z en Afabilidad.

g) ***Verdadera***. El puntaje Z correspondiente es – 1,5 (una desviación típica y media por debajo de la media del grupo).

PRÁCTICA V

Uno de los coeficientes que se utiliza para determinar la existencia y el grado de relación lineal entre dos variables es el ***coeficiente de correlación de Pearson***.

El coeficiente de correlación de Pearson, representado por la letra ***r***, es un índice que mide la relación lineal entre dos variables aleatorias cuantitativas.

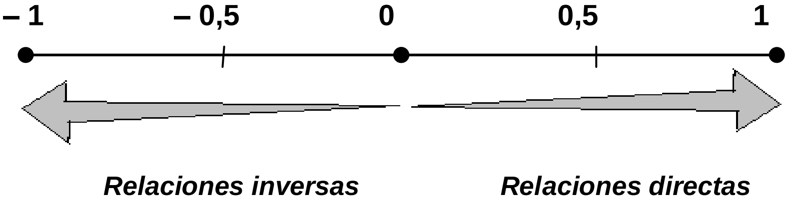
La fórmula que lo define es:

$$r_{xy} = \frac{S_{xy}}{S_x \cdot S_y} = \frac{n \cdot \sum (x_i \cdot y_i) - \sum x_i \cdot \sum y_i}{\sqrt{n \cdot \sum x_i^2 - \left(\sum x_i\right)^2} \cdot \sqrt{n \cdot \sum y_i^2 - \left(\sum y_i\right)^2}}$$

En la interpretación de una correlación hay que separar dos aspectos distintos:

- *Cuantía*: se refiere al grado en que la relación entre dos variables queda bien descrita con un índice de asociación lineal como ***r***.
- *Sentido*: se refiere al tipo de relación.

Recordemos que una de las propiedades de este coeficiente es que no puede valer más de 1 ni menos de – 1.



Cuanto más cercano quede un coeficiente del valor cero (relación baja o nula), menos apto es el modelo lineal como descripción de la relación entre las variables.

EJERCICIO 1

Indicar cuáles de las siguientes afirmaciones sobre el coeficiente de correlación lineal r de Pearson son siempre verdaderas y cuáles son siempre falsas y cuáles contingentes.

a) Está comprendido entre -1 y 1.
b) Es adimensional.
c) Si es 0, las variables son independientes.
d) Si r = 0,07 para las variables peso y longitud de un objeto expresados en kg y cm respectivamente el valor del coeficiente pasa a ser r = 0,7 para las mismas variables expresadas en gramos y metros.
e) La correlación lineal entre el tiempo empleado en completar una tarea y el número de errores cometidos en la misma es r = - 0,6. De ello se deduce que una de las causas que contribuyen importantemente en el número de errores cometidos en la tarea es el tiempo que se emplea en resolverla.

a) ***Siempre verdadera***.

Por la propiedad que hemos mencionado anteriormente.

b) ***Siempre verdadera***.

Recordemos que la covarianza de dos variables presentaba el problema de carecer de valores máximo y mínimo estables, comunes a todos los casos, que permitan su interpretación directa. El coeficiente de correlación de Pearson dio solución a este problema ya que consiste en hallar también un promedio de productos cruzados, pero no de las puntuaciones diferenciales, sino de las puntuaciones típicas. Por eso a veces se dice que la correlación es una *covarianza adimensional*.

c) ***Contingente***

Pueden ser independientes o puede haber una relación que no se aproxima a la linealidad. Por ejemplo, la relación entre activación y rendimiento que adopta la forma de una U invertida (el rendimiento máximo se obtiene con niveles medios de activación, mientras que con niveles bajos o demasiados altos el rendimiento disminuye).

d) ***Siempre falsa***.

El coeficiente *r* no depende de la unidad de medición por basarse su fórmula, precisamente, en las puntuaciones típicas y no en las puntuaciones diferenciales.

e) ***Contingente***

Si hay un buen diseño de la experiencia puede darse una interpretación causal. Una correlación no sólo puede aparecer entre una variable-causa y una variable-efecto, sino también entre variables entre las que no hay relaciones de causalidad, por lo tanto, no deben inferirse relaciones de causalidad *sólo* a partir del hecho de haber obtenido un valor de ***r*** sustancialmente diferente de cero.

EJERCICIO 2

En la siguiente tabla se informan los valores de r y las características de los diagramas de dispersión en cinco situaciones. Establezca la correspondencia correcta entre ambas columnas.

Valores de r	Los puntos del diagrama de dispersión…
a) 1	1) tienden a ubicarse en torno a una recta creciente.
b) - 1	2) están distribuidos en el plano sin ninguna estructura particular.
c) 0	3) tienden a ubicarse en torno a una recta decreciente.
d) 0,9	4) están perfectamente alineados en una recta creciente.
e) - 0,7	5) están perfectamente alineados en una recta decreciente.

Para establecer la correspondencia entre ambas columnas recordemos que:

- En una relación lineal directa, los valores altos de Y tienden a emparejarse con valores altos de X, los valores intermedios de Y con valores intermedios de X, y los valores bajos de Y con valores bajos de X, con lo cual, los pares de valores (X,Y) que conforman el diagrama de dispersión se ubican en torno a una recta creciente.
- En una relación lineal inversa, los valores altos de Y tienden a emparejarse con valores bajos de X, los valores intermedios de Y con valores intermedios de X, y los valores bajos de Y con valores altos de X, con lo cual, los pares de valores (X,Y) que conforman el diagrama de dispersión se ubican en torno a una recta decreciente.

a)	4)
b)	5)
c)	2)
d)	1)
e)	3)

EJERCICIO 3

En la siguiente tabla se informan los valores de r y los pares de variables estudiadas en cinco situaciones. Establezca la correspondencia entre ambas columnas que podría esperarse suponiendo que los diseños de investigación fueron adecuados.

Valores de r	Variables estudiadas
a) - 0,4	1) Puntaje en ansiedad medido con el test A y con el test B.
b) 0,8	2) Estatura de ciertos sujetos y su puntaje en razonamiento verbal.
c) 0,9	3) Puntaje en un ítem dicotómico de un test de habilidades matemáticas y el puntaje en el resto del test.
d) 0,02	4) Puntaje en un test de Bienestar y puntaje en un test de Estrés.
e) 0,3	5) Puntaje en los ítems pares y puntaje en los ítems impares de un test de analogías verbales.

1) El puntaje en ansiedad medido con diferentes tipos de test dará tendencias similares para cada sujeto, valores altos de ansiedad medidos con el test A se emparejan con valores altos de ansiedad medidos con el test B, valores intermedios en el test A con valores intermedios en el test B y valores bajos en el test A con valores también bajos en el test B, con lo cual, la relación es directa. Tanto los ítems de un test como del otro miden el mismo rasgo, cabe esperar una correlación alta y positiva (**b**).
2) No hay relación entre la estatura de un sujeto y su puntaje en razonamiento verbal, ***r*** es aproximadamente 0 (**d**).
3) Se trata de una prueba para discriminar a los sujetos de más alto rendimiento de los de más bajo rendimiento. Si el ítem tiene buen poder discriminatorio, es de esperar que los sujetos de alto rendimiento tiendan a contestar bien y los de bajo rendimiento, mal. Como el ítem es dicotómico no es de esperar una representación gráfica lineal en el diagrama de dispersión, pero sí una correlación positiva (**e**).

4) Se trata de dos test diferentes que miden el Bienestar y el Estrés de un sujeto. En el diagrama de dispersión, cabe esperar una correlación negativa (**a**). Se supone que los puntajes altos en el test de Bienestar se emparejarán con puntajes bajos en el test de Estrés, los puntajes intermedios de un test con puntajes intermedios del otro, y los puntajes bajos en el test de Bienestar con altos en el test de Estrés.

5) Todos los ítems del test miden el mismo rasgo. Se trata de evaluar la consistencia interna del test, con lo cual, si la consistencia interna es buena, cabe esperar una correlación alta y positiva (**c**).

Cabe mencionar, respecto de los puntos 1) y 5), que, si bien en ambos casos se trata de una correlación alta y positiva, se espera más homogeneidad en un mismo test (test de analogías verbales) que en dos test diferentes (test A y test B), aunque midan el mismo rasgo.

a)	4)
b)	1)
c)	5)
d)	2)
e)	3)

PRÁCTICA VI

EJERCITACIÓN

DISTRIBUCIÓN BINOMIAL

La distribución binomial es un modelo para variables cuantitativas discretas. Recordemos que para que una distribución de probabilidad de una variable se ajuste al modelo binomial se deben cumplir una serie de requisitos.

El primero de ellos es que se base en una *variable Bernoulli*. Una variable Bernoulli es aquella que sólo admite dos valores. Generalmente, estos valores son: 1, si se cumple una cierta condición (llamaremos *éxito* al cumplimiento de la misma), y 0 si no se cumple (llamaremos *fracaso* al no cumplimiento de la condición). La probabilidad asociada al éxito se la denota con la letra ***p***.

El segundo requisito es que haya una repetición de ***n*** ensayos de la variable Bernoulli en los que la probabilidad de éxito, es decir, la probabilidad de que se verifique la condición especificada, sea *constante*. Los resultados de cada ensayo son *independientes* entre sí, es decir, el conocer el resultado de una observación no aumenta ni disminuye la probabilidad de obtener éxito en otra observación.

Por último se define una variable, **X**, que representa el número de casos que, en la secuencia de los *n* ensayos dicotómicos, verifican la condición especificada.

Decimos entonces, que la variable X se distribuye binomialmente con parámetros *n* y *p*. Esto lo representamos de la siguiente forma:

$$X \sim B\ (n\ ;\ p)$$

Los valores que toma la variable binomial X oscilan entre 0 (la condición especificada no se verifica en ningún ensayo) y *n* (la condición especificada se verifica en todos los ensayos).

Veamos los siguientes ejemplos.

Ejemplo 1

Si de una población de 500 esquizofrénicos, de los que un 30 % sufren alteración cerebral, extraemos una muestra aleatoria simple de 17 sujetos, ¿cuál es la probabilidad de que haya más de 8 sujetos con alteración cerebral en la muestra? (Botella 13.4.2 – Pág. 337)

En primer lugar, identificamos los parámetros *n* y *p*.

Se trata de seleccionar una muestra aleatoria simple de 17 sujetos. Consideramos la selección de cada uno como un ensayo, por lo tanto, ***n* = 17**.

Debemos verificar si cada sujeto cumple o no la condición de sufrir una alteración cerebral. Cada vez que seleccionamos un individuo de la población son posibles sólo dos resultados: el individuo sufre de alteración cerebral o no. Llamemos *éxito* a la condición de, seleccionado un individuo, presente alteración cerebral, y *fracaso*, a la condición en que el sujeto no presente alteración cerebral. Sabemos que el 30% de la población sufre una alteración cerebral, es decir, 30 de cada 100 sujetos presentan este problema. Esto significa que la probabilidad de encontrar, en la población de 500 esquizofrénicos, uno que verifique la condición especificada es $\frac{30}{100} = 0{,}30$, ***p* = 0,30**.

La probabilidad de éxito, *p*, es siempre la misma en cada ensayo. También existe independencia entre los resultados de cada uno ya que, el saber si un sujeto posee alteración cerebral no aumenta ni disminuye la probabilidad de que otro sujeto también lo tenga.

Definimos la variable **X**:

X = “Cantidad de esquizofrénicos con alteración cerebral”

La variable X cuenta la cantidad de éxitos en los 17 ensayos. Así, la variable X se distribuye binomialmente con parámetros *n* = 17 y *p* = 0,30.

X ~ B (17; 0,30)

Nos interesa calcular la probabilidad de que haya más de 8 sujetos con alteración cerebral en la muestra, es decir que X debe tomar un valor mayor que 8. Lo expresamos de la siguiente forma:

P (X > 8)

Las probabilidades asociadas a valores de variables binomiales se encuentran tabuladas para algunos valores de *n* y de *p*. Lo único que debemos hacer es buscar adecuadamente en la tabla.

$$P(X > 8) = P(X = 9) + P(X = 10) + P(X = 11) + P(X = 12) + P(X = 13) + P(X = 14) +$$
$$+ P(X = 15) + P(X = 16) + P(X = 17)$$
$$P(X > 8) = 0{,}028 + 0{,}009 + 0{,}003 + 0{,}001 = 0{,}041$$

Respuesta: **0,041**

Ejemplo 2

Sabiendo por experiencias previas que el 70 % de los alumnos de Psicología de la UBA promociona Estadística y suponiendo que esta proporción se mantenga el presente cuatrimestre, calcule la probabilidad de que en una muestra aleatoria de 10 alumnos:

a) Por lo menos uno no promocione
b) No promocionen cinco
c) Promocione por lo menos la mitad
d) No promocione a lo sumo dos
e) Promocionen más de cuatro y menos de ocho

Debemos seleccionar una muestra aleatoria de 10 alumnos, por lo tanto, ***n* = 10**.

Se nos pide calcular probabilidades asociadas a la *promoción* y a la *no promoción* de la asignatura Estadística.

Consideremos un éxito, el seleccionar un alumno de Psicología de la UBA que promocione Estadística. Cada vez que seleccionamos un alumno, sólo hay dos resultados posibles, promociona o no promociona Estadística. Los resultados son mutuamente excluyentes. De acuerdo a experiencias previas, el 70 % de los alumnos promociona Estadística, esto significa que 70 de cada 100 alumnos promociona la materia, la probabilidad de éxito, entonces, es 0,70, ***p* = 0,70**.

Se cumple la condición de estabilidad (la probabilidad de éxito es constante en las 10 observaciones) y la condición de independencia (la probabilidad de éxito en una observación no aumenta ni disminuye si conocemos el resultado de otra observación), Definimos la variable X:

X = “Cantidad de alumnos de Psicología de la UBA que *promociona* Estadística”

La variable X se distribuye binomialmente con parámetros *n* = 10 y *p* = 0,70.

X ~ B (10; 0,70)

a) Por lo menos un alumno no promocione la materia significa que promocionen como máximo 9 alumnos. Este cálculo es un poco tedioso, ya que equivale a buscar todas las probabilidades desde X = 0 hasta X = 9 y sumar dichos valores.

$$P(X \le 9) = P(X = 0) + P(X = 1) + P(X = 2) + \ldots\ldots\ldots\ldots\ldots + P(X = 9)$$

Hay una forma más simple de hallar este valor. Sabemos que la suma de las probabilidades asignadas al conjunto de valores de la variable es 1.

$$\underbrace{P(X = 0) + P(X = 1) + P(X = 2) + \ldots\ldots\ldots\ldots\ldots + P(X = 9)}_{P(X \le 9)} + P(X = 10) = 1$$

Por lo tanto:

$$P(X \le 9) = 1 - P(X = 10) = 1 - 0{,}028 = \underline{\mathbf{0{,}972}}$$

b) Que no promocionen 5 de los 10 alumnos es equivalente a decir que promocionen los 5 alumnos restantes.

$$P(X = 5) = \underline{\mathbf{0{,}103}}$$

c) Calcular la probabilidad de que promocione <u>*por lo menos*</u> la mitad, significa calcular la probabilidad de que promocionen 5 o más alumnos:

$$P(X \ge 5) = P(X = 5) + P(X = 6) + P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10)$$

= 0,103 + 0,2 + 0,267 + 0,233 + 0,121 + 0,028 = **<u>0,952</u>**

d) Hallar la probabilidad de que no promocionen <u>*a lo sumo*</u> dos alumnos, es equivalente a que promocionen <u>*por lo menos*</u> ocho.

$$P(X \ge 8) = P(X = 8) + P(X = 9) + P(X = 10)$$

= 0,233 + 0,121 + 0,028 = **<u>0,382</u>**

e) Para calcular la probabilidad de que promocionen más de cuatro y menos de ocho,sumamos las probabilidades de que la variable X tome los valores 5, 6 y 7:

$$P(4 < X < 8) = P(X = 5) + P(X = 6) + P(X = 7)$$

= 0,103 + 0,2 + 0,267 = **<u>0,57</u>**

A veces, no resulta fácil expresar, por ejemplo, las probabilidades correspondientes a la *no promoción* de la materia, en términos de la *promoción* de la misma. En estos casos podemos proceder, también, de la siguiente forma:

Definimos la variable X como sigue:

X = “Cantidad de alumnos de Psicología de la UBA que ***no promociona*** Estadística”

Ahora, el éxito será seleccionar un alumno de Psicología de la UBA que *no* promocione Estadística. Si el 70% promociona la materia, el 30% no promociona: ***p* = 0,3**.

La variable X se distribuye normalmente con parámetros *n* = 10 y *p* = 0,3.

X ~ B (10; 0,30)

a) Se desea calcular la probabilidad de que <u>por lo menos</u> uno no promocione, es decir, uno o más no promocionen Estadística.

$$P(X \geq 1) = 1 - P(X = 0) = 1 - 0{,}028 = \underline{\mathbf{0{,}972}}$$

b) Probabilidad de que no promocionen cinco: $P(X = 5) = \underline{\mathbf{0{,}103}}$

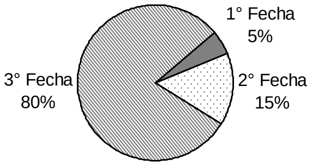
d) Hallar la probabilidad de que no promocionen <u>a lo sumo</u> dos alumnos, es lo mismo que decir que no promocionen <u>como máximo</u> dos.

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

$$= 0{,}028 + 0{,}121 + 0{,}233 = \underline{\mathbf{0{,}382}}$$

EJERCICIO 1

Estadísticas recientes de nuestra facultad afirman que de cada 10 inscriptos para rendir finales en diciembre sólo 3 se presentan. El siguiente gráfico reproduce cómo se distribuyen los alumnos que se presentan en las tres fechas tentativas:



A partir de la información brindada

I. calcule la probabilidad de que de una muestra aleatoria de 12 alumnos inscriptos

a) más de 4 y menos de 8 no se presenten a rendir.

b) por lo menos 4 se presenten.

II. calcule la probabilidad de que de una muestra aleatoria de 6 alumnos que se presentan

a) a lo sumo 1 no se presente en la tercera fecha.

b) ninguno se presente en la primera fecha.

I. Definimos la variable X:

X = “Cantidad de alumnos, de la Facultad de Psicología de la UBA, inscriptos que se presentan para rendir finales en diciembre en la asignatura Estadística”

La muestra está constituida por 12 alumnos (***n* = 12**); los 12 alumnos son seleccionados de los alumnos que se inscriben para rendir examen final. Para cada alumno seleccionado tenemos dos resultados posibles: que se presente para rendir final o no. Son resultados mutuamente excluyentes. Consideremos un éxito la condición de seleccionar un alumno inscripto que se presente para rendir examen final. Sabemos que 3 de cada 10 inscriptos se presentan, por lo tanto, la probabilidad de éxito es 0,3 (***p* = 0,3**).

La variable X se distribuye binomialmente con parámetros *n* = 12 y *p* = 0,3.

X ~ B (12; 0,3)

a) Más de 4 y menos de 8 no se presenten a rendir, significa que 5, 6 o 7 alumnos no se presenten a rendir, y esto es equivalente a que se presenten 7, 6 o 5 alumnos respectivamente (que no se presenten 5 es equivalente a que se presenten 7, que no se presenten 6 es equivalente a que se presenten 6 y que no se presenten 7 es equivalente a que se presenten 5).

Luego, tenemos que calcular la probabilidad de que X tome un valor comprendido entre 5 y 7 inclusive:

$$P(5 \leq X \leq 7) = P(X=5) + P(X=6) + P(X=7) = 0{,}158 + 0{,}079 + 0{,}029 = \underline{\mathbf{0{,}266}}$$

b) Por lo menos 4 se presenten significa que se deben presentar como mínimo 4 alumnos:

$$P(X \geq 4) = 1 - P(X < 4) = 1 - [P(X=0) + P(X=1) + P(X=2) + P(X=3)] =$$

$$= 1 – [0{,}014 + 0{,}071 + 0{,}168 + 0{,}240] = 1 – 0{,}493 = \underline{\mathbf{0{,}507}}$$

II. Se selecciona, ahora, una muestra de 6 alumnos, (***n* = 6**), pero sobre los alumnos inscriptos que se han presentado a rendir examen. En este caso, seleccionado un alumno, el mismo puede presentarse en una y sólo una de las tres fechas. Si consideramos éxito el presentarse a una de ellas y fracaso el presentarse a las dos fechas restantes, la variable queda dicotomizada artificialmente.

a) Sea X = “Cantidad de alumnos inscriptos que se presentan en la tercera fecha”

Llamemos éxito a la condición de que un alumno inscripto se presente en la tercera fecha de examen. En la tercera fecha se presenta el 80% de los alumnos inscriptos, ***p* = 0,8**.

X ~ B (6; 0,8)

A lo sumo 1 no se presente es equivalente a que se presenten como mínimo 5 alumnos:

$$P(X \geq 5) = P(X=5) + P(X=6) = 0{,}393 + 0{,}262 = \underline{\mathbf{0{,}655}}$$

b) Sea X = “Cantidad de alumnos inscriptos que se presentan en la primera fecha”

Llamemos éxito a la condición de que un alumno inscripto que se presenta en la primera fecha de examen. En la primera fecha se presenta el 5% de los alumnos inscriptos, por lo tanto la probabilidad de éxito es 0,05, ***p* = 0,05**

X ~ B(6; 0,05)

$$P(X=0) = \underline{\mathbf{0{,}735}}$$

EJERCICIO 2

Un grupo de 4 estudiantes que han preparado juntos el final de Psicoanálisis Freud II decidieron analizar la posibilidad que tienen de aprobarlo. Recurrieron a la encuesta que publica la cátedra en su página de Internet y hallaron los siguientes datos:

En diciembre	Se presentaron	Aprobaron	Desaprobaron
1998	654	591	63
1999	699	643	56
2000	729	670	59
2001	619	578	41

Ellos están interesados en saber cuál es la probabilidad de que

a) apruebe todo el grupo.
b) alguno desapruebe.

Estime estas probabilidades para responder a sus inquietudes.

En las cuatro fechas publicadas en internet se presentó un total de 2701 alumnos. De ellos, aprobaron un total de 2482 alumnos y desaprobaron un total de 219.

El grupo está formado por 4 estudiantes (***n* = 4**). Cada alumno que se presentó al final de Psicoanálisis Freud II, aprobó o desaprobó la materia. Llamemos éxito a la condición de aprobar el examen. Si, según las fechas publicadas, aprobaron 2482 de un total de 2701 alumnos, la probabilidad de éxito es $\frac{2482}{2701} \approx 0{,}9$, ***p* = 0,9**

Sea X = "Cantidad de alumnos que aprueban Psicoanálisis Freud II",

$$X \sim B(4;0{,}9)$$

a) P (X = 4) = **<u>0,656</u>**

b) Alguno desapruebe es equivalente a que por lo menos uno desapruebe o aprueben como máximo 3 de los 4 estudiantes:

$$P(X \leq 3) = 1 - P(X = 4) = 1 - 0{,}656 = \underline{\mathbf{0{,}344}}$$

EJERCICIO 3

Una prueba de percepción visual consiste en responder a cuál de los dos matices, previamente presentados, corresponde el color verde que aparece sucesivamente en 12 estampados diferentes. El resultado de la prueba se considera "muy satisfactorio" cuando se ha respondido acertadamente en, por lo menos, 9 de las 12 presentaciones.

En prácticas previas un sujeto ha logrado acertar el 80% de las representaciones. ¿Cuál es la probabilidad de que al realizar la prueba

a) obtenga el resultado "muy satisfactorio"?
b) no obtenga un resultado "muy satisfactorio" suponiendo que contestó completamente al azar en todos los casos?
c) Si se hicieran sólo 8 presentaciones, ¿cuál debería ser el punto de corte para clasificar un resultado como "muy satisfactorio" si se desea que la probabilidad de obtenerlo contestando por azar sea pequeña, digamos inferior a 0,05?

Se presentan al sujeto 12 estampados diferentes; consideramos cada presentación como un ensayo, frente al cual el sujeto debe responder a cuál de los dos matices corresponde el color verde. Por lo tanto, ***n* = 12**.

En cada ensayo la respuesta del sujeto puede ser correcta o incorrecta. Llamemos éxito a la condición de acertar en cada ensayo el matiz correspondiente. Si el sujeto, en prácticas previas, logró acertar el 80% de las representaciones, significa que acertó 8 de cada 10 presentaciones, ***p* = 0,8**.

Sea X = "Cantidad de respuestas correctas en las 12 presentaciones"

$$X \sim B(12;\ 0{,}8)$$

a) Obtiene el resultado "muy satisfactorio" si acierta en, por lo menos, 9 de las 12 presentaciones,

$$P(X \geq 9) = P(X = 9) + P(X = 10) + P(X = 11) + P(X = 12) = 0{,}236 + 0{,}283 + 0{,}206 + 0{,}069$$

= **<u>0,794</u>**

b) Si contestó al azar, tiene las mismas posibilidades de responder correctamente como de responder en forma incorrecta, ***p* = 0,5**.

Sea X = "Cantidad de respuestas correctas, habiendo respondido en las 12 presentaciones por azar"

X ~ B(12; 0,5)

No obtiene un resultado "muy satisfactorio" si contesta acertadamente en menos de 9 presentaciones.

$$P(X < 9) = 1 - P(X \geq 9) = 1 - [P(X = 9) + P(X = 10) + P(X = 11) + P(X = 12)]$$

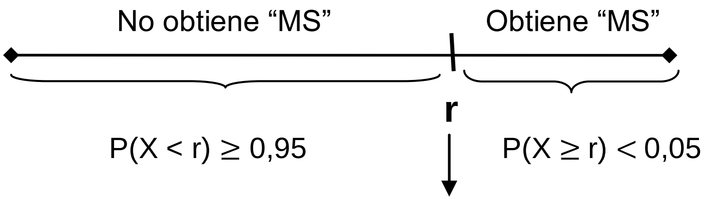
= 1 – [0,054 + 0,016 + 0,003 + 0] = **<u>0,927</u>**

c) Se hacen sólo 8 presentaciones, ***n* = 8**; el sujeto responde por azar, ***p* = 0,5**.

Sea X = "Cantidad de respuestas correctas habiendo contestado en las 8 presentaciones por azar'.

X ~ B(8; 0,5)

Llamemos **r** al punto de corte. El valor **r** representa la cantidad de respuestas correctas que exigimos debe dar el sujeto para obtener el resultado "muy satisfactorio" contestando por azar. Deseamos que la probabilidad de obtener este resultado sea inferior a 0,05, es decir, $P(X \geq r) < 0{,}05$. Esto es equivalente a que la probabilidad de que el sujeto *no obtenga* el resultado "muy satisfactorio", debe ser 0,95, y esto ocurre cuando da una cantidad de respuestas correctas inferior a **r**, es decir, $P(X < r) \geq 0{,}95$.



Cantidad mínima de respuestas correctas que debe dar el sujeto para obtener "muy satisfactorio"

Reproduzcamos la parte de la tabla que nos interesa:

		P
n	**x**	**0,5**
8	0	0,004
	1	0,031
	2	0,109
	3	0,219
	4	0,273
	5	0,219
	6	0,109
	7	0,031
	8	0,004

La suma de las probabilidades desde 0 hasta 6 es **0,964**

Debemos buscar el valor de **r** tal que la suma de las probabilidades desde 0 hasta dicho valor se excede 0,95. Esto ocurre para X = 6. Por lo tanto, debemos exigir que un sujeto, respondiendo por azar, para obtener el resultado “muy satisfactorio” debe responder correctamente, como mínimo en 7 de las presentaciones.

Respuesta: **r = 7**

EJERCICIO 4 (Datos ficticios)

En cierta región de una provincia argentina Ignacio (sociólogo) y Lucía (maestra) están estudiando el fenómeno de la deserción escolar a nivel primario. De una encuesta del último año a varias escuelas obtuvieron la siguiente información:

Causas de deserción	Nro. de casos
Necesidad de trabajar	180
Distancia a la escuela	40
Enfermedad	55
Dificultades intelectuales	20
Otros	5

Habiendo iniciado el año lectivo en dichas escuelas un total de 6000 alumnos.

Ignacio se pregunta qué probabilidad tendrá de que entre 12 niños que abandonen la escuela el año entrante, más de la tercera parte lo haga debido a la necesidad de trabajar. Lucía se pregunta con qué probabilidad concluirá el año conservando a sus 15 alumnos.

Suponiendo que el año entrante tenga un comportamiento similar al anterior con respecto al fenómeno de la deserción, responda a

a) Ignacio.
b) Lucía.

a) Para responder a Ignacio debemos conocer el total de niños que abandonaron la Escuela el año anterior. Sumamos el número de casos que han abandonado por diferentes motivos, esto es 300 niños.

A Ignacio le interesa conocer la probabilidad que tendrá entre 12 niños (***n* = 12**) que abandonen la escuela el año entrante, más de la tercera parte lo haga debido a la necesidad de trabajar, es decir, más de 4, lo haga por este motivo.

Nuevamente estamos ante una variable que no es dicotómica naturalmente, pero podemos dicotomizarla artificialmente. Cada niño que abandona la escuela puede hacerlo por necesidad de trabajar o por otra causa. Llamemos éxito a la condición de que un niño abandone la escuela por necesidad de trabajar, de 300 niños, 180 lo hacen por esta causa. La probabilidad de éxito, entonces, es $\frac{180}{300} = 0{,}6$, ***p* = 0,6**.

Sea X = “Cantidad de niños que, entre los que abandonan la escuela, lo hagan por necesidad de trabajar”

X ~ B(12; 0,6)

$$P(X > 4) = 1 - P(X \leq 4) = 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)]$$

$$= 1 - [0{,}002 + 0{,}012 + 0{,}042] = 1 - 0{,}056 = \underline{\mathbf{0{,}944}}$$

b) Lucía se pregunta con qué probabilidad concluirá el año conservando a sus 15 alumnos (***n* = 15**).

Sea X = “Cantidad de niños que abandonan la escuela”

De un total de 6000 alumnos, 300 abandonaron la escuela. Cada niño abandona la escuela o permanece en ella. Considerando un éxito la condición de que un niño abandone la escuela (no importa ahora la causa del abandono), la probabilidad de éxito es $\frac{300}{6000} = 0{,}05$, ***p* = 0,05**.

X ~ B(15; 0,05)

Conservar a sus 15 alumnos es equivalente a ninguno abandone la escuela, es decir que X debe ser igual a 0.

$P(X = 0) =$ **0,463**

EJERCICIO 5 (Datos ficticios)

Resultados de la selección para los integrantes de la Nueva Policía Bonaerense

2 de cada 5 postulantes para integrar la Policía tienen trastornos psicológicos

Se trata de 1500 aspirantes que perdieron la posibilidad de integrarse a las fuerzas policiales por presentar un nivel de agresividad muy elevado y falta de sensibilidad social. La mayoría son hombres y casi la mitad pertenece a clase baja. El **80% afirmó que ser policía es su verdadera vocación**. A su vez, **450 de ellos declararon que su ingreso promedio es inferior al que ganarían como policía**.

A partir de la información brindada:
I. calcule la probabilidad de que de una muestra aleatoria de 6 aspirantes a integrar la nueva policía bonaerense
  a) por lo menos 5 tengan trastornos psicológicos.
  b) más de la mitad no hayan presentado trastorno psicológicos.

II. calcule la probabilidad de que de una muestra aleatoria de 5 postulantes con trastornos psicológicos
  a) a lo sumo 3 no elijan esta profesión por vocación.
  b) ninguno posea un ingreso inferior al que aspira.

I. Se trata de una muestra aleatoria de 6 aspirantes, ***n* = 6**.

Sea X = “Cantidad de aspirantes a integrar la nueva policía bonaerense que sufren un trastorno psicológico”.

Consideremos un éxito la condición de ser un aspirante que tenga un trastorno psicológico. Si 2 de cada 5 aspirantes tiene este problema, la probabilidad de éxito es $\frac{2}{5} = 0{,}4$, ***p* = 0,4**.

X ~ B(6; 0,4)

a) Que por lo menos 5 tengan trastornos psicológico, significa que 5 o más aspirantes sufran este problema.

$P(X \geq 5) = P(X = 5) + P(X = 6) = 0{,}037 + 0{,}004 = \underline{\mathbf{0{,}041}}$

b) Más de la mitad significa que 4 o más aspirantes no presenten trastornos psicológicos, y esto es equivalente a que sí presenten algún trastorno un máximo de 2 aspirantes:

$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0{,}047 + 0{,}187 + 0{,}311 = \underline{\mathbf{0{,}545}}$

II. Trabajamos, ahora, con una muestra de 5 postulantes con trastornos psicológicos, ***n* = 5**.

a) Sea X = "Cantidad de postulantes con trastornos psicológicos que dicen que ser policía es su verdadera vocación"

Consideremos éxito a la condición de ser un aspirante con trastornos psicológicos que expresa que ser policía es su verdadera vocación. El 80% afirmó que ser policía era su verdadera vocación, por lo tanto, la probabilidad de éxito es 0,8, ***p* = 0,8**.

X ~ B(5; 0,8)

A lo sumo 3 no elijan esta profesión por vocación es equivalente a que sí elijan la profesión por este motivo como mínimo 2 de los aspirantes:

$P(X \geq 2) = P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) = 0{,}051 + 0{,}205 + 0{,}410 + 0{,}328 =$

= **<u>0,994</u>**

b) Sea X = "Cantidad de aspirantes con trastornos psicológicos cuyo ingreso promedio es inferior al que ganarían como policía"

De los 1500 aspirantes, 450 declararon que su ingreso promedio es inferior al que ganarían como policía. Si consideramos un éxito la condición de tener un ingreso promedio inferior al que gana un policía, la probabilidad de éxito es $\frac{450}{1500} = 0{,}3$, ***p* = 0,3**.

X ~ B(5; 0,3)

Ninguno posea un ingreso inferior al que aspira significa que la variable X tome valor 0:

$P(X = 0) = \underline{\mathbf{0{,}168}}$

EJERCICIO 6

Gastón sabe que el 95% de las personas logran responder correctamente una pregunta de aritmética de un test de inteligencia. Pero interroga a sus colegas acerca de la cantidad de tiempo que tardaron aquellos pacientes que acertaron con su respuesta

Respuesta Correcta	Consulta 1	Consulta 2	Consulta 3
Antes de 30seg.	34	18	44
Después de 30seg.	83	42	98

I. Suponiendo que Gastón administra el test a 14 personas calcule la probabilidad de que: a) alguno responda incorrectamente.

b) por lo menos 13 respondan correctamente.

II. Suponiendo que 16 personas respondan correctamente calcule la probabilidad de que: a) más del 75% de ellos lo hagan después de 30 segundos.

b) ninguno lo haga antes de 30 segundos.

I. El test se aplica a 14 personas, ***n* = 14**. Consideremos éxito a la condición de responder correctamente a la pregunta de aritmética. Si el 95% de las personas lo hace correctamente, la probabilidad de éxito es 0,95, ***p* = 0,95**.

Sea X = “Cantidad de personas que responden correctamente a la pregunta de aritmética de un test de inteligencia”

X ~ B(14; 0,95)

a) Que alguno responda incorrectamente una pregunta es equivalente a que respondan correctamente un máximo de 13 personas:

$$P(X \leq 13) = 1 - P(X = 14)1 - 0{,}488 = \underline{\mathbf{0{,}512}}$$

b) Por lo menos 13 respondan correctamente es equivalente a que 13 o más personas lo hagan:

$$P(X \geq 13) = P(X = 13) + P(X = 14) = 0{,}359 + 0{,}488 = \underline{\mathbf{0{,}847}}$$

II. Seleccionamos una muestra de 16 personas que responden correctamente a esta pregunta, ***n* = 16**.

a) El 75% de 16 es 12. En cada ensayo hay sólo dos resultados posibles, cada persona responde correctamente antes de 30 segundos o después de 30 segundos. Consideremos éxito la condición de responder después de los 30 segundos. De las 319 personas, 223 respondieron después de 30 segundos, por lo tanto, la probabilidad de éxito es $\frac{223}{319} \approx 0{,}7$.

Sea X = “Cantidad de personas que responden correctamente a la pregunta de aritmética después de los 30 segundos”.

X ~ B(16; 0,7)

$$P(X > 12) = P(X = 13) + P(X = 14) + P(X = 15) + P(X = 16)$$

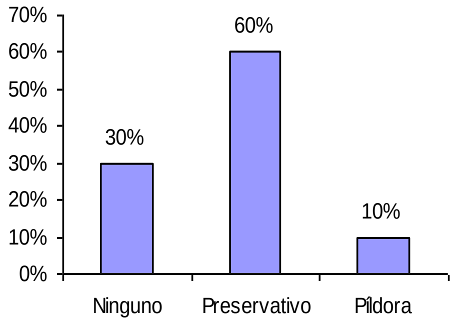
$$= 0{,}146 + 0{,}073 + 0{,}023 + 0{,}003 = \underline{\mathbf{0{,}245}}$$

b) Hallar la probabilidad de que ninguno lo haga antes de los 30 segundos es equivalente a que todos contesten correctamente después de los 30 segundos.

$$P(X = 16) = \underline{\mathbf{0{,}003}}$$

EJERCICIO 7 (Datos ficticios)

El caso de un joven de 14 años que se convirtió en padre motivó la publicación de una estadística aseguró que 1 de cada 5 adolescentes se inician sexualmente antes de cumplir 15 años. Consultados por el método anticonceptivo empleado por ellos se ha podido confeccionar el gráfico



A partir de la información brindada

I. calcule la probabilidad de que de una muestra aleatoria de 9 adolescentes

a) menos de 6 hayan iniciado sexualmente antes de los 15 años.

b) la mayoría no se haya iniciado sexualmente antes de cumplir los 15 años.

II. calcule la probabilidad de que de una muestra aleatoria de 5 jóvenes que se hayan iniciado antes de los 15 años

a) a lo sumo 4 hayan usado algún método anticonceptivo.

b) el 40% haya usado preservativo.

I. Seleccionamos una muestra aleatoria de 9 adolescentes, ***n* = 9**. De acuerdo a la estadística, 1 de cada 5 adolescentes se inician sexualmente antes de cumplir 15 años. Considerando éxito la condición mencionada, la probabilidad de seleccionar un adolescente que se inicie sexualmente antes de cumplir 15 años es 0,2, ***p* = 0,2**.

Sea X = “Cantidad de adolescentes que se inician sexualmente antes de cumplir 15 años”

X ~ B(9; 0,2)

a) $P(X<6)=P(X=0)+P(X=1)+P(X=2)+P(X=3)+P(X=4)+P(X=5)$

= 0,134 + 0,302 + 0,302 + 0,176 + 0,066 + 0,017 = **<u>0,997</u>**

b) La mayoría, es decir, 5 o más adolescentes, no se haya iniciado sexualmente antes de los 15 años es equivalente a que la minoría, 4 o menos, se haya iniciado sexualmente antes de los 15 años:

$P(X\leq 4)=P(X=0)+P(X=1)+P(X=2)+P(X=3)+P(X=4)$

= 0,134 + 0,302 + 0,302 + 0,176 + 0,066 = **<u>0,980</u>**

II. Se selecciona una muestra aleatoria de 5 jóvenes, ***n* = 5**.

a) Sea X = “Cantidad de adolescentes que, habiéndose iniciado sexualmente antes de los 15 años, usa algún método anticonceptivo”.

Según la estadística publicada, el 60% usa preservativo y el 10%, píldora. Si consideramos éxito la condición de que un adolescente use algún método anticonceptivo, la probabilidad de éxito es 0,7, ***p* = 0,7**.

X ~ B(5; 0,7)

$P(X\leq 4)=P(X=0)+P(X=1)+P(X=2)+P(X=3)+P(X=4)$

= 0,002 + 0,028 + 0,132 + 0,309 + 0,360 = **<u>0,831</u>**

b) Sea X = “Cantidad de adolescentes que, habiéndose iniciado sexualmente antes de los 15 años, usa preservativo”.

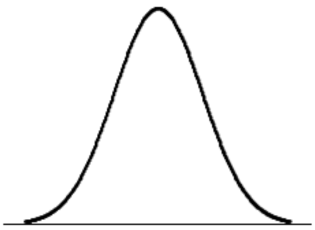
Considerando éxito la condición de que un adolescente use preservativo, la probabilidad de éxito es 0,6, ***p* = 0,6**.

X ~ B(5; 0,6)

El 40% de 5 es 2: $P(X=2)=$ **<u>0,230</u>**

DISTRIBUCIÓN NORMAL

La distribución normal es un modelo para variables cuantitativas continuas. Muchas de las variables estudiadas por los psicólogos presentan la forma de esta distribución.



Esta curva es la representación gráfica de una función matemática que nos permite conocer la probabilidad de encontrar una observación cualquiera, si conocemos en cuántas desviaciones típicas se aparta de la media de su distribución. Debido a que la forma de la curva normal es estándar, existe un porcentaje conocido de valores por debajo y por encima en cualquier punto en particular.

Recordemos que el área total bajo la curva normal es igual a 1 y contiene al 100% de todos los valores de la variable, de forma que la *población total* se expresa en el gráfico como el *área total bajo la curva normal*.

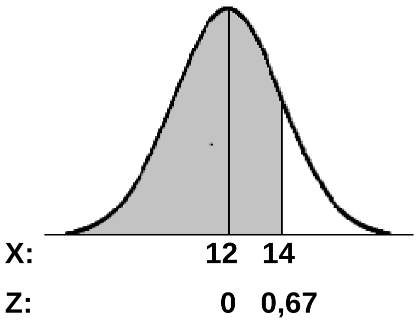
EJERCICIO 8

Suponiendo que los puntajes en una prueba de razonamiento verbal se distribuyen normalmente con media 12 y desvío estándar 3 en un grupo normativo, calcule:

a) El rango percentilar de un sujeto que obtiene 14 puntos.
b) El puntaje correspondiente al sujeto que es superado por el 90% del grupo.
c) El centil 95.
d) La probabilidad de que en una muestra aleatoria de 10 sujetos, la mayoría tenga un puntaje superior a 14,52.

Al resolver problemas que implican áreas que se encuentran bajo la curva normal, es conveniente realizar la gráfica de dicha curva, ubicar los datos del problema y localizar cuál es el área correspondiente al problema.

a) Un sujeto que obtiene 14 puntos en la prueba de RV se encuentra por encima de la media del grupo normativo, 12. Representemos gráficamente esta situación.



El área sombreada contiene todas las calificaciones menores que 14. Para determinar el rango percentilar de 14, primero debemos convertir el puntaje 14 en su puntaje z correspondiente:

$$z = \frac{14-12}{3} = \frac{2}{3}0,\widehat{6} \quad \Rightarrow \quad z = 0,67$$

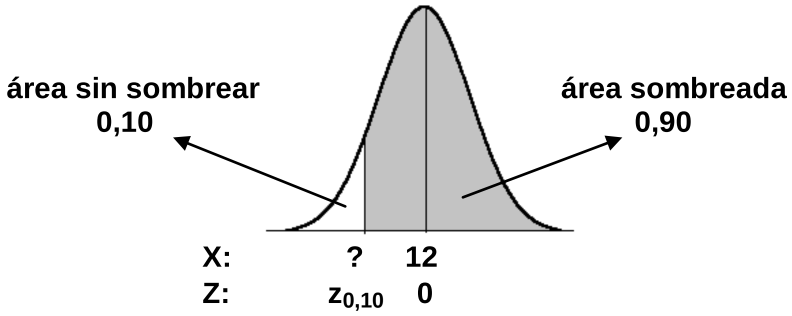
Recurrimos, ahora, a la tabla de áreas bajo la distribución normal. Esta tabla contiene tres columnas tituladas como **z**, **área** y **ordenada**. Las mismas corresponden, respectivamente, a los valores de las puntuaciones z, el área bajo la curva y a la *izquierda* de cada puntaje z, y a la altura de la curva normal en cada valor de z. La tabla nos informa acerca de la *probabilidad* de obtener una puntuación cualquiera por encima o por debajo de cualquier puntuación típica o entre dos puntuaciones típicas.

La puntuación típica, correspondiente al puntaje 14 de la prueba de RV, obtenida es 0,67. Localizamos el valor 0,67 en la columna de z y leemos la entrada correspondiente al área; este valor es 0,7486. El mismo corresponde al área bajo la curva y a la izquierda de z, y representa la probabilidad de encontrar en el grupo normativo una puntuación menor o igual a 14.

El rango percentilar de 14 lo obtenemos multiplicando el valor de área encontrado por 100:

0,7486 · 100 = **74,86**

b) También, a veces, conocemos el área y queremos determinar el dato correspondiente. En este caso, queremos encontrar el puntaje que es superado por el 90% del grupo; esto es equivalente a decir que el área bajo la curva y a la derecha de este puntaje desconocido es 0,90 o, lo que es lo mismo, el área bajo la curva y a la izquierda de ese valor es 0,10 (recordemos que el *área total bajo la curva es 1* y que la tabla de la normal nos da los valores correspondientes al área a la *izquierda* de z). Representamos gráficamente esta situación:



Este problema es justo el inverso del anterior. Conocida el área determinamos, en primer lugar cuál es el valor de z correspondiente, es decir, cuál es el valor de z que deja a la izquierda un área igual a 0,10. Para ello, en la tabla, buscamos en la columna de área el valor 0,1000 o el que más se aproxime a él; el puntaje z correspondiente a esta área es -1,28.

En símbolos esto se expresa de la siguiente manera:

$$\mathbf{Z_{0,10} = -1,28}$$

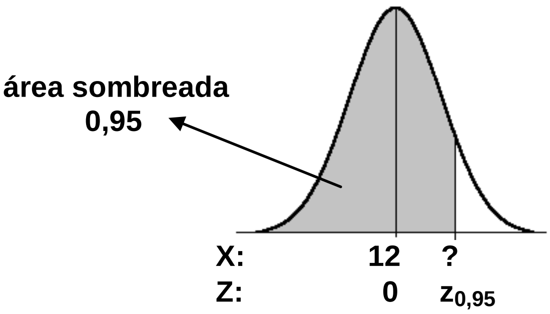
Para determinar el dato en bruto, reemplazamos los valores correspondientes en la ecuación (1) y despejamos la incógnita *x*:

$$z = \frac{x-\mu}{\sigma} \qquad (1)$$

$$-1,28 = \frac{x-12}{3} \quad \Rightarrow \quad -1,28 \cdot 3 + 12 = x \quad \Rightarrow \quad x = 8,16$$

X = 8,16

c) El centil 95 es la puntuación que deja por debajo de sí el 95% de las observaciones y es superado por el 5% de las mismas. En términos de área bajo la curva normal, debemos buscar la puntuación que deja a la izquierda un área igual a 0,95 y a la derecha un área igual a 0,05.



Nuevamente buscamos en la tabla, en la columna de área, el valor más próximo a 0,9500; en este caso, encontramos dos valores equidistantes de 0,9500, ellos son 0,9495 y 0,9505; tomamos el punto medio de los respectivos valores de z, ese valor es 1,645. Reemplazando en la fórmula anterior obtenemos la puntuación buscada:

$$z = \frac{x-\mu}{\sigma} \qquad (1)$$

$$z_{0,95} = 1{,}645 \quad \Rightarrow \quad 1{,}645 = \frac{x-12}{3} \quad \Rightarrow \quad 1{,}645 \cdot 3 + 12 = x \quad \Rightarrow \quad x = 16{,}935$$

$\underline{\mathbf{C_{95} = 16,94}}$

d) En este problema se combinan los dos modelos: binomial y normal.

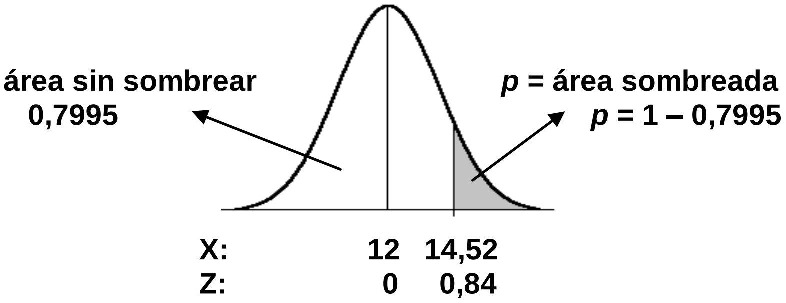
Se selecciona una muestra aleatoria de 10 sujetos, ***n* = 10**. Cada sujeto seleccionado obtiene un puntaje superior a 14,52 o no (tenemos sólo dos resultados posibles). Consideramos éxito la condición de superar dicho puntaje.

Sea X = “Cantidad de sujetos que obtienen en la prueba de razonamiento verbal un puntaje superior a 14,52”

La variable X se distribuye binomialmente con parámetros *n* = 10 y probabilidad de éxito *p*:

$$\mathbf{X \sim B(10;\ \boldsymbol{p})}$$

La probabilidad de éxito, ***p***, es desconocida, pero sabemos que los puntajes de la prueba de RV se distribuyen normalmente con media 12 y desvío estándar 3; por lo tanto, encontrar la probabilidad de que un sujeto obtenga un puntaje superior a 14,52, equivale a encontrar el área bajo la curva de la normal a la derecha del puntaje z correspondiente. Dicha área es el valor ***p*** buscado.



$$z = \frac{14{,}52 - 12}{3} = 0{,}84$$

Buscamos en la tabla el área correspondiente al valor de z hallado. Dicha área es 0,7995. Recordemos que la tabla nos informa sobre el área a izquierda de z. Como el área total bajo la curva es 1, el área a la derecha del puntaje 14,52 será la que se obtiene de restar a 1 el valor 0,7995:

$$1 - 0{,}7995 = 0{,}2005 \quad \Rightarrow \quad p = 0{,}2$$

Ahora estamos en condiciones de hallar la probabilidad de que en una muestra de 10 sujetos, la mayoría tenga un puntaje superior a 14,52.

X ~ B(10; *0,2*)

Buscamos en la tabla de la binomial P(X > 5):

$$P(X > 5) = P(X = 6) + P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10)$$

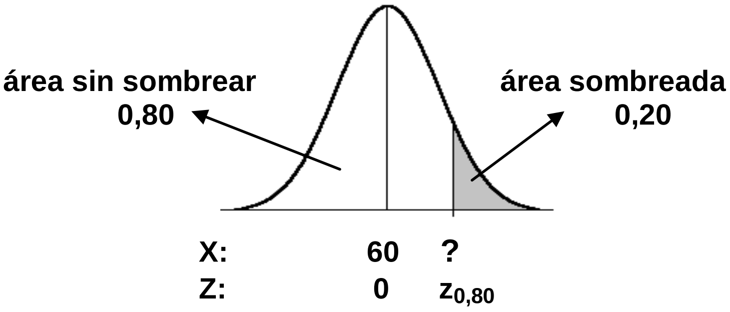
= 0,006 + 0,001 = **<u>0,007</u>**

EJERCICIO 9

Para ingresar a cierta universidad se debe rendir un examen de ingreso y sólo hay cupo para el 20% de los aspirantes. Suponiendo que los puntajes se distribuyen normalmente en el grupo de aspirantes con media 60 y desviación estándar 15.

a) ¿Qué nota mínima de aprobación debe fijarse?
b) ¿Cuál es la probabilidad de que en una muestra aleatoria de 15 aspirantes más de la tercera parte aprueben el examen de ingreso?

a) La nota mínima de aprobación corresponde al C_{80}, es decir, la puntuación que supera al 80% de las observaciones y es superada por el 20% de las mismas. En términos de áreas, es la puntuación que deja a su izquierda un área igual a 0,80 y a su derecha un área igual a 0,20. El esquema gráfico que representa esta situación es el siguiente:



$$z_{0,80} = 0{,}84 \quad \Rightarrow \quad 0{,}84 = \frac{x - 60}{15} \quad \Rightarrow \quad 0{,}84 \cdot 15 + 60 = x \quad \Rightarrow \quad x = 72{,}6$$

<u>C_{80} = 72,6</u>

b) En este punto, nuevamente, se combina el modelo binomial con el modelo normal.

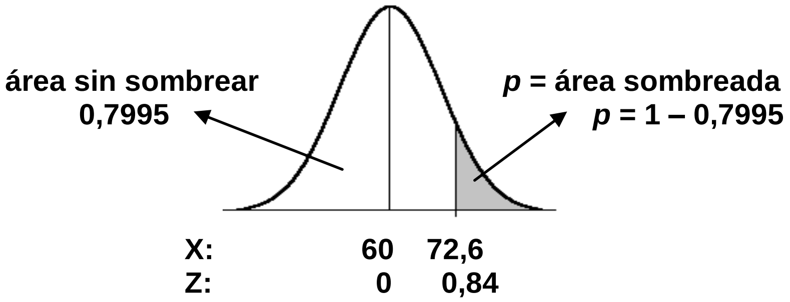
Se selecciona una muestra aleatoria de 15 aspirantes a ingresar en la universidad, ***n* = 15**. Cada aspirante seleccionado obtiene como mínimo un puntaje de 72,6 o un puntaje inferior (tenemos sólo dos resultados posibles). Consideramos éxito la condición de obtener un puntaje igual o mayor a 72,6.

Sea X = “Cantidad de sujetos que obtienen en el examen de ingreso un puntaje igual o mayor a 72,6”

La variable X se distribuye binomialmente con parámetros *n* = 15 y probabilidad de éxito *p*:

X ~ B(15; *p*)

La probabilidad de éxito, ***p***, es desconocida, pero sabemos que los puntajes del examen de ingreso se distribuyen normalmente con media 60 y desvío estándar 15; por lo tanto, encontrar la probabilidad de que un sujeto obtenga un puntaje igual o superior a 72,6, equivale a encontrar el área bajo la curva de la normal a la derecha del puntaje z correspondiente. Dicha área es el valor ***p*** buscado.



$$z = \frac{72{,}6 - 60}{15} = 0{,}84$$

Buscamos en la tabla el área correspondiente al valor de z hallado. Dicha área es 0,7995. Esta es el área a izquierda de z; el área a la derecha del puntaje 72,6 se obtiene restando de 1 el valor hallado:

1 – 0,7995 = 0,2005 ⇒ *p* = 0,2

Hallamos la probabilidad de que en una muestra de 15 aspirantes, más de la tercera parte, es decir, más de 5 aspirantes, aprueben el examen de ingreso, esto es, que obtengan un puntaje igual o mayor que 72,6.

X ~ B(15; *0,2*)

Buscamos en la tabla de la binomial P(X > 5):

$$P(X > 5) = 1 - P(X \leq 5) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)$$

= 1- [0,035 + 0,132 + 0,231 + 0,250 + 0,188 + 0,103 = 1 – 0,939 = **0,061**

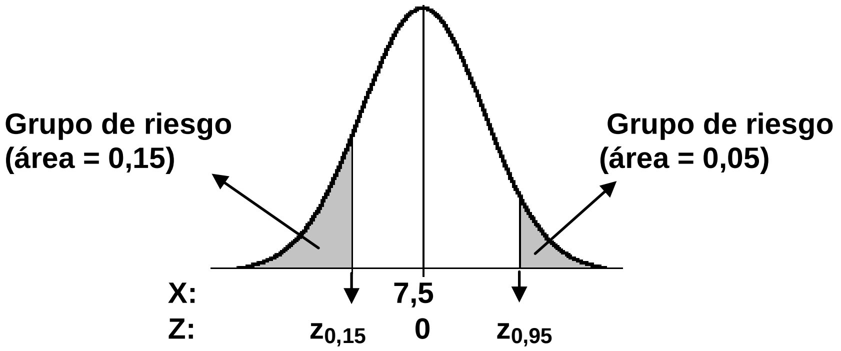
EJERCICIO 10

La medicina tradicional considera normal dormir 7,5 horas con un desvío de 0,5 horas para mujeres de 40 años. Una cantidad de horas inferior al centil 15 y superior al centil 95 se considera riesgosa para este grupo poblacional por considerarse el insomnio y la hipersomnia dos indicadores claros de trastornos depresivos. Sobre una base de 160 mujeres estudiadas responder:

a) ¿Cuáles son las cantidades máxima y mínima de horas de sueño que debe dormir una mujer para pertenecer al grupo de riesgo?

b) Si una paciente duerme 8 horas ¿a cuántas personas supera?
c) ¿Cuántas horas duerme una mujer ubicada en el último centil?
d) ¿Cuál es la probabilidad de que una mujer duerma entre 7 y 8 horas?

a) El grupo de riesgo está formado por las mujeres de 40 años que duermen una cantidad de horas inferior al centil 15 o superior al centil 95. Por lo tanto, debemos encontrar cuáles son las puntuaciones que corresponden a dichos centiles. El centil 15 es la puntuación que deja por debajo de sí el 15% de las observaciones, en términos de áreas, deja a izquierda un área igual a 0,1500. El centil 95 es la puntuación que supera al 95% de las observaciones, deja a izquierda un área igual a 0,9500.



En la tabla, en la columna de área, buscamos los valores que más se aproximen a 0,1500 y 0,9500:

$$z_{0,15} = -1{,}04 \quad \Rightarrow \quad -1{,}04 = \frac{x-7{,}5}{0{,}5} \quad \Rightarrow \quad -1{,}04 \cdot 0{,}5 + 7{,}5 = x \quad \Rightarrow \quad x = 6{,}98$$

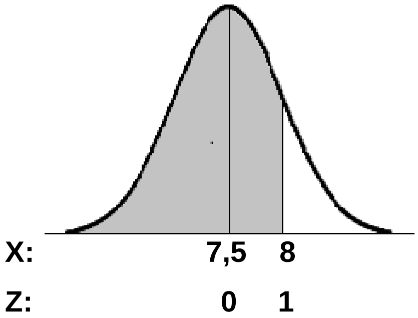
C_{15} = 6,98

$$z_{0,95} = 1{,}645 \quad \Rightarrow \quad 1{,}645 = \frac{x-7{,}5}{0{,}5} \quad \Rightarrow \quad 1{,}645 \cdot 0{,}5 + 7{,}5 = x \quad \Rightarrow \quad x = 8{,}3225$$

C_{95} = 8,32

Las cantidades máxima y mínima de horas de sueño que debe dormir una mujer para pertenecer al grupo de riesgo son 6,98 y 8,32.

b) El porcentaje de pacientes que supera una persona que duerme 8 horas se corresponde con el área que deja a izquierda su puntaje z

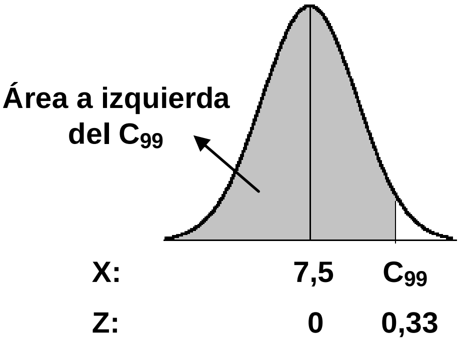


$$z = \frac{8-7{,}5}{0{,}5} = 1 \quad \Rightarrow \quad z = 1$$

Buscamos en la tabla el área a la izquierda de z = 1, dicha área es 0,8413. Una persona que duerme 8 horas supera al 84,13% de las obervaciones.

El 84,13% de 160 mujeres es **134**.

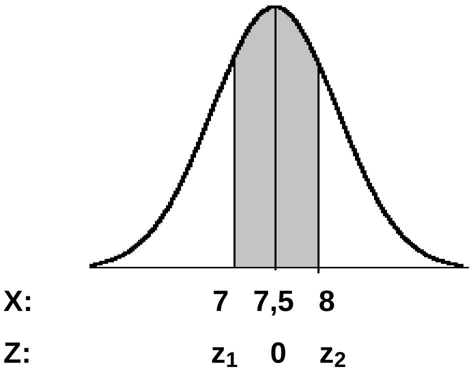
c) El último centil es el C_{99}, putuación que supera al 99% de las observaciones y es superada por el 1% de las mismas. Buscamos en la tabla, en la columna de área, la más cercana a 0,9900. El valor de z correspondiente es 2,33.



$$z_{0,99} = 2,33 \quad \Rightarrow \quad 2,33 = \frac{x-7,5}{0,5} \quad \Rightarrow \quad 2,33 \cdot 0,5 + 7,5 = x \quad \Rightarrow \quad x = 8,665$$

C_{99} = 8,67

d) En este caso debemos calcular primero dos puntajes z, el correspondiente a 7 y el correspondiente a 8, pues debemos hallar la probabilidad de que una persona duerma entre 7 y 8 horas; esto equivale a hallar el área encerrada entre estos dos valores.



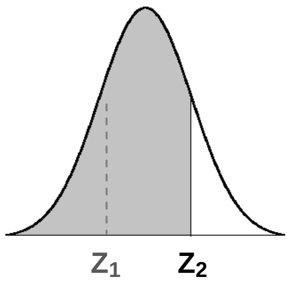
$$z_1 = \frac{7-7,5}{0,5} = -1 \qquad\qquad z_2 = \frac{8-7,5}{0,5} = 1$$

Necesitamos calcular el área encerrada por ambos valores. Para ello buscamos en la tabla el área que cada uno de los puntajes z hallados deja a su izquierda.

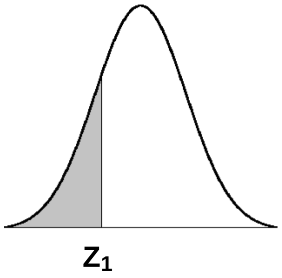
El área a izquierda de z_1 es 0,1587, y el área a izquierda de z_2 es 0,8413. El área de la región comprendida entre ambos valores se obtiene restando el área a izquierda de z_1 del área a izquierda de z_2.

0,8413 – 0,1587 = **0,6826**

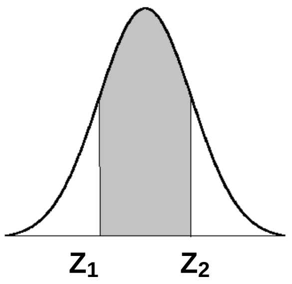
Gráficamente, esto es:



Área a izquierda de z_2



Área a izquierda de z_1



Diferencia entre las áreas

0,8413 - 0,1587 = **<u>0,6826</u>**

La probabilidad de que una mujer duerma entre 7 y 8 horas es 0,6826.

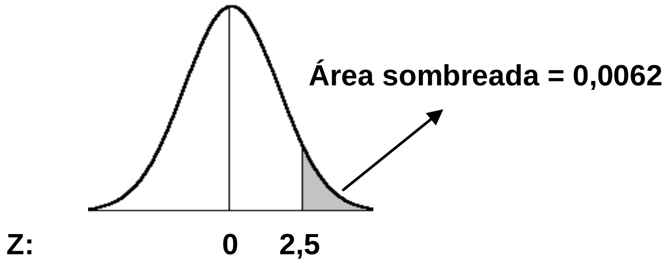
EJERCICIO 11

Los puntajes de un test que mide el grado de expresión del enojo de un sujeto se distribuye normalmente con media 25 y desvío 2 en una población argentina. La teoría califica como “irritable” a las personas que están a menos de un desvío y medio por encima de la media y como “violento” a las personas que exceden dos desvíos y medio por encima de la media. Sobre una base de 120 personas evaluadas responda:

a) ¿Qué porcentaje de sujetos cabe esperar en el grupo “violento”?
b) ¿Cuántos sujetos cabe esperar en el grupo “irritable”?
c) ¿Cuál es el puntaje correspondiente al tercer decil?

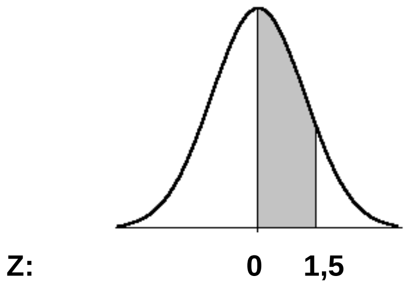
a) El grupo “violento” está formado por las personas que se encuentran a más de dos desvíos y medio por encima de la media, es decir, las personas que tienen un puntaje z mayor a 2,5. El área a izquierda de z es 0,9938, por lo tanto el área que corresponde al grupo violento es:

$$1 – 0,9938 = 0,0062$$



Multiplicando el área por 100 obtenemos el porcentaje buscado.
El porcentaje de sujetos que cabe esperar en el grupo “violentos” es: **<u>0,62%.</u>**

b) El grupo “irritable” es el que está a menos de un desvío y medio por encima de la media.

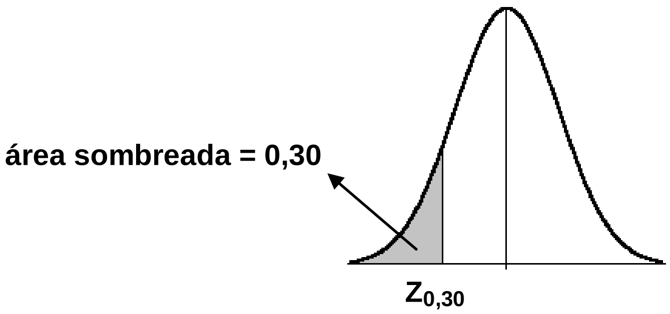


El área comprendido entre ambos valores se encuentra, igual que anteriormente, restando las áreas respectivas. El valor 0 corresponde a la media de las puntuaciones z; siendo la curva simétrica respecto de este valor, el área a la izquierda de 0 es 0,5000. El área a izquierda de 1,5 es 0,9332.

0,9332 – 0,5000 = 0,4338 ⇒ el 43,38% de sujetos cabe esperar en el grupo “irritable”

El 43,38% de 120 es: **52 personas**.

c) El D_3 es la puntuación que supera el 30% de las observaciones y es superada por el 70% de las mismas. Gráficamente:



Buscamos en la tabla el área que más se aproxima a 0,3000: $z_{0,30} = -0{,}52$

$$z_{0,30} = -0{,}52 \quad \Rightarrow \quad -0{,}52 = \frac{x-25}{2} \quad \Rightarrow \quad -0{,}52 \cdot 2 + 25 = x \quad \Rightarrow \quad x = 23{,}96$$

<u>D_3 = 23,96</u>

EJERCICIO 12

El estrés ocupacional percibido por los 200 empleados que realizan atención de cliente en una empresa fue medido con el Inventario de Estrés Ocupacional (OSI) de Osipow y Spokane. Esta variable se distribuye normalmente y utilizan los puntajes T para establecer el nivel de estrés.

Nivel de Estrés	Referencias
Fuerte	T > 69,5
Leve	59,5 < T < 69,5
Normal	39,5 < T < 59,5
Ausencia	T < 39,5

Calcular:

a) ¿En cuántos empleados cabe esperar que estén ausentes los indicadores de estrés?

b) ¿Qué porcentaje de sujetos percibirán un leve nivel de estrés?

c) ¿Qué puntuaciones como mínimo y como máximo será necesario establecer si se desea seleccionar al 50% de los sujetos que obtienen las puntuaciones centrales?
d) ¿Qué nivel de estrés percibe un sujeto que es superado por el 60% de la población?

Recordemos que la construcción de una escala derivada parte de unas puntuaciones directas, éstas se tipifican, y después se transforman linealmente en otras puntuaciones equivalentes a las puntuaciones originales (Ver Práctica IV de esta guía).

Las puntuaciones T tienen media 50 y desviación típica 10.

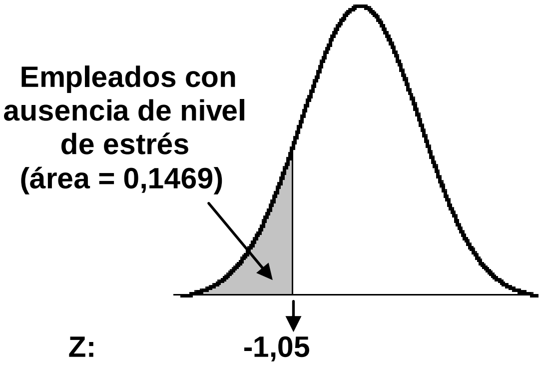
Así, la transformación en puntajes T es: $T_i = 10 \cdot z_i + 50$

Para responder a los interrogantes del ejercicio, tenemos que recorrer el camino en sentido inverso.

Si $T_i = 10 \cdot z_i + 50$ entonces $z_i = \frac{T_i - 50}{10}$

a) Los indicadores de estrés están ausentes en aquellos empleados que obtuvieron un puntaje T < 39,5. Debemos hallar la probabilidad de encontrar un sujeto que haya obtenido tal puntaje:

$$P(T < 39{,}5) = P\left(z < \frac{39{,}5 - 50}{10}\right) = P(Z < -1{,}05) = 0{,}1469$$



Multiplicando el área por 100 obtenemos el porcentaje de empleados en los que cabe esperar que estén ausentes los indicadores de estrés, este porcentaje es 14,69%.

El 14,69% de 200 es: 0,1469 · 200 = 29,38

Respuesta: **29 empleados**

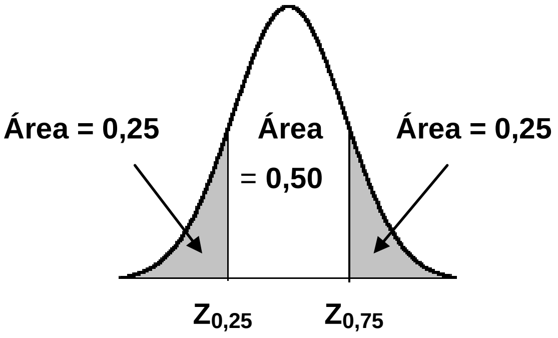
b) Los empleados que perciben un leve nivel de estrés obtuvieron un puntaje T comprendido entre 59,5 y 69,5.

$$P(59{,}5 < T < 69{,}5) = P\left(\frac{59{,}5 - 50}{10} z < \frac{69{,}5 - 50}{10}\right) = P(0{,}95 < Z < 1{,}95) =$$

$$P(z < 1{,}95) - P(z < 0{,}95) = 0{,}9744 - 0{,}8289 = 0{,}1455$$

El porcentaje de empleados que perciben un leve nivel estrés se obtiene multiplicando el área hallada por 100: 0,1455 · 100 = **14,55**

c) Las puntuaciones que acotan el 50% central del área total son aquéllas que dejan a su izquierda y a su derecha, áreas iguales a 0,25:

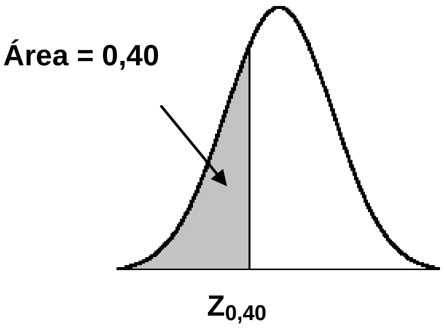


$z_{0,25}$ = - 0,67 $\Rightarrow$ T = 10 · (- 0,67) + 50 = 43,3

$z_{0,25}$ = 0,67 $\Rightarrow$ T = 10 · 0,67 + 50 = 56,7

Respuesta: **<u>43,3 y 56,7</u>**

d) Debemos hallar el puntaje correspondiente al C_{40}, puntuación que supera al 40% de las observaciones y es superada por el 60% de las mismas; por lo tanto, este puntaje es el que deja a izquierda un área igual a 0,40.



$z_{0,40}$ = - 0,25 $\Rightarrow$ T = 10 · (- 0,25) + 50 = 47,5

Respuesta: **<u>47,5</u>**

EJERCICIO 13

De los 200 alumnos que rindieron un examen para obtener una beca, sólo 44 han podido recibirla por superar los 65 puntos. Si las calificaciones de estos alumnos están distribuidas normalmente con una desviación típica de 6 puntos ¿cuál es el promedio de esta población?

Los puntajes correspondientes a las calificaciones se distribuyen normalmente; podemos usar el modelo normal para hallar probabilidades y áreas correspondientes a los mismos.

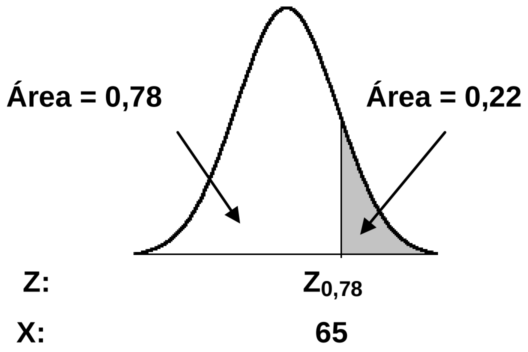
Los 44 alumnos que recibieron la beca representan el 22% de los 200 que rindieron el examen, pues:

200 ———————— 100%

44 ———————— x

$$x = \frac{44 \cdot 100}{200} = 22\%$$

El 22% se corresponde con un área igual a 0,22, y representa a los alumnos que obtuvieron un puntaje superior a 65 puntos. El mismo puntaje es la puntuación que supera al 78% de las observaciones, es decir, deja un área a izquierda igual a 0,78:



Buscamos, en la tabla, el puntaje z correspondiente a 65, entrando por la columna de área; dicho puntaje es 0,77.

Para hallar el promedio de la población hacemos uso de la fórmula que permite transformar puntuaciones directas en puntuaciones típicas:

$$z_i = \frac{x_i - \mu}{\sigma} \quad \Rightarrow \quad 0,77 = \frac{65 - \mu}{6} \quad \Rightarrow \quad 0,77 \cdot 6 - 65 = -\mu \quad \Rightarrow \quad -60,38 = -\mu \quad \Rightarrow \mu = 60,38$$

El promedio es 60,38

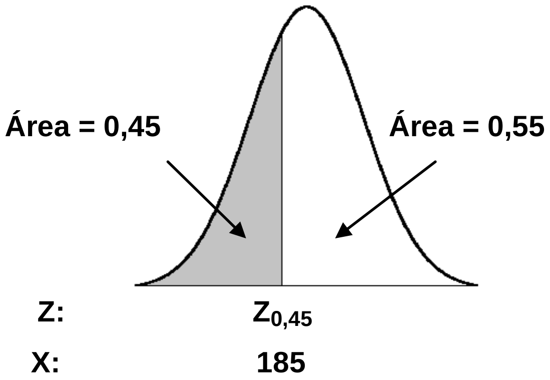
Nota: Recordemos que podemos hacer uso del programa Excel para obtener tanto los valores correspondientes a los puntajes z como para obtener las áreas respectivas. En este caso, el valor de z es 0.77219321, más aproximado que el que obtuvimos con la tabla. Reemplazando este valor en la fórmula obtendríamos para µ un valor aproximadamente igual a **60,368**.

EJERCICIO 14

El tiempo necesario para concluir con el armado de un rompecabezas de 100 piezas se distribuye normalmente con media 210 minutos. Sabiendo que el 45% de los sujetos tarda como máximo 185 minutos ¿cuál es la desviación estándar de esta distribución?

El procedimiento es similar al anterior, la diferencia es que ahora, la incógnita es el desvío estándar.

El puntaje 185 supera al 45% de las observaciones. Haciendo uso del modelo normal, esto es equivalente a decir que es el puntaje que deja a izquierda un área igual a 0,45.



Nuevamente, entrando por la columna de área, buscamos el puntaje z que le corresponde a 185; dicho puntaje es – 0,13:

$$Z_{0,45} = -0,13$$

$$z_i = \frac{x_i - \mu}{\sigma} \quad \Rightarrow \quad -0,13 = \frac{185-210}{\sigma} \quad \Rightarrow \quad -0,13 \cdot \sigma = -25 \quad \Rightarrow \quad \sigma = 192,31$$

Respuesta: **<u>192,31</u>**

EJERCICIO 15

Una prueba de rendimiento consiste en 50 ítemes de 4 opciones de las cuales sólo una es verdadera. El requerimiento para aprobar es contestar correctamente por lo menos el 40% de la prueba, esto es, 20 ítemes.

En este tipo de pruebas interesa conocer qué probabilidad tiene de aprobar un alumno que contesta todo por azar.

Si definimos la variable X = Cantidad de respuestas acertadas entre las 50 habiendo contestado todas por azar

a) ¿Qué modelo de distribución de probabilidades es adecuado para la variable X? Justifique.
b) Calcule la probabilidad de aprobar por azar.
c) Halle la nota mínima de aprobación para que apruebe por azar a lo sumo el 2% de los alumnos.

a) Por la manera en que la variable está definida y por las características de la experiencia el modelo más adecuado es el binomial de parámetros n = 50 y p = 0,25. En efecto, contestar todos los ítemes de la prueba por azar puede pensarse como repetir n = 50 veces un ensayo de Bernoulli, ya que es contestar un ítem donde sólo caben dos posibilidades: responderlo acertadamente (éxito) o incorrectamente (fracaso). Al responder completamente al azar los ítemes resultan independientes, ya que todos tienen la misma cantidad de opciones (cuatro) y sólo una es verdadera. Por tanto, se cumplen las dos condiciones de éxito constante en todos los ensayos.

b) Como se aprueba con 20 ítemes o más respuestas correctas debemos calcular P(X ≥ 20). La probabilidad exacta es la suma de las probabilidades binomiales desde 20 hasta 50:

$$P(X \geq 20) = P(X = 20) + P(X = 21) + P(X = 22) + P(X = 23) + \ldots + P(X = 50)$$

Sumar tantos términos es tedioso y actualmente hay calculadoras que dan directamente el resultado de tales sumas. También hay tablas de probabilidades “acumuladas” que dan esta información. Pero para quien no tenga estos recursos puede apelar al uso de la distribución normal, la cual da una buena aproximación de las probabilidades binomiales acumuladas cuando se cumplen ciertas condiciones de simetría que no detallaremos aquí. Llamaremos X_N a la variable normalmente distribuida que usaremos para aproximar las probabilidades de la variable binomial X. Sus parámetros serán la media y desviación estándar de X, o sea:

Media de $X = n.p = 50.\frac{1}{4} = 12,5$ = Media de X_N

Desviación estándar de $X = \sqrt{n.p.(1-p)} = \sqrt{50.\frac{1}{4}.\frac{3}{4}} = 3,06$ = Desvío estándar de X_N

Resumiendo:

$$X_N \sim N\ (12{,}5\ ;\ 3{,}06)$$

Para aproximar la probabilidad binomial por la normal debe hacerse una corrección que consiste en considerar media unidad más; es decir que

$$P(X \geq 20) = P\ (X_N \geq 19{,}5)$$

Esto se llama corrección por continuidad y es necesaria cuando a una variable que es discreta se la trata como continua. (El lector interesado en una mayor explicación puede remitirse al cuadro 13.4, pág. 325 de Botella)

$$P\ (X_N \geq 19{,}5) = P\left[\frac{X_N - 12{,}5}{3{,}06} \geq \frac{19{,}5 - 12{,}5}{3{,}06}\right] = P\ (\ Z \geq 2{,}29\ ) = 0{,}011 \quad \text{donde } Z \sim N\ (0;1)$$

Este resultado puede interpretarse así: de cada 100 alumnos que resuelvan toda la prueba por azar, aproximadamente uno solo aprobará.

c) Debemos obtener un puntaje x de manera tal que:

$$P\ (X_N \geq x\ ) < 0{,}02$$

Estandarizando tenemos que:

$$P\ (X_N \geq x\ ) = P\left[Z \geq \frac{x - 12{,}5}{3{,}06}\right] < 0{,}02$$

Buscando en la tabla normal estándar el valor z que acumula a izquierda una probabilidad de 0,98 hallamos que, aproximadamente:

$$Z > 2{,}06 \quad \text{o sea que} \quad \left[Z \geq \frac{x - 12{,}5}{3{,}06}\right] > 2{,}06$$

y despejando obtenemos

$$x > 3{,}06\ .\ 2{,}06 + 12{,}5 > 18{,}8036$$

y como la cantidad mínima para aprobar de respuestas correctas es un número entero, habrá que fijar este mínimo en:

PRÁCTICA VII

EJERCITACIÓN

CONTRASTE DE HIPÓTESIS SOBRE UNA MEDIA

El contraste de hipótesis sobre una media sirve para tomar decisiones respecto del verdadero valor poblacional que corresponde a la media de una variable. Nos encontramos por lo tanto, ante un diseño con una muestra.

Si extraemos una muestra aleatoria de tamaño *n* de una población normal N(μ,σ) y calculamos la media $\overline{X}$, la misma es un estadístico distribuido normalmente $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$.

Sabemos también, por el teorema central del límite, que, aun desconociendo la forma de la población de donde extraemos la muestra, el estadístico $\overline{X}$ tiende a distribuirse normalmente $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$, a medida que el tamaño *n* de esta última va aumentando.

Estandarizando $\overline{X}$:

$$E = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$, tiene, aproximadamente, distribución normal estándar.

Los supuestos o condiciones de aplicabilidad son:

- La varianza poblacional σ^2 es conocida.
- La variable es normal o el tamaño de la muestra es suficientemente grande.

Nota: En las pruebas unilaterales para la diferencia de medias de datos pareados o independientes, tanto el signo en la hipótesis alternativa > (derecha) o < (izquierda) como el signo del valor observado del estadístico de contraste (positivo o negativo), dependen del sentido en que se haya planteado la diferencia, el cual es arbitrario. Para las respuestas a este tipo de problemas se eligió una de las dos posibles formulaciones pero lo importante es que todo esté en coherencia.

EJERCICIO 1

En una investigación para validar la versión en español del Test Conductual de Memoria de Rivermead (RBMT) para población mayor de 70 años, Alonso, M. y Prieto, P. (2004) informan una media de 16,1 con una desviación estándar de 4,2 para las personas entre 70 y 75 años de un grupo normativo de Tenerife. Si en una muestra aleatoria de 81 personas del mismo rango de edad de la ciudad A se observara una media de 16,8 puntos

a) ¿habría suficiente evidencia al 5% de que el rendimiento en el test de las personas entre 70 y 75 años de la ciudad A es en promedio superior al del grupo normativo de Tenerife?

b) ¿Qué valores de la media muestral harían la diferencia estadísticamente significativa en la dirección de la hipótesis planteada?

Nota. Los valores de la media y desviación estándar están redondeados a un decimal. El artículo completo puede hallarse en http://www.psicothema.com/pdf/1201.pdf

Al resolver un contraste de hipótesis trabajamos con datos referidos a poblaciones y muestras. Es importante diferenciar correctamente cuáles son los datos que corresponden a la población y cuáles los que corresponden a la muestra, como así también la simbología utilizada en los mismos.

En nuestro ejercicio disponemos de los siguientes datos:

Desvío conocido →

DATOS DE LA POBLACIÓN	DATOS DE LA MUESTRA
$\mu = 16,1$	$\bar{x} = 16{,}8$
$\sigma = 4,2$	n = 81

a) Debemos responder si la evidencia empírica de la que disponemos es suficiente para afirmar que el rendimiento en el test de las personas entre 70 y 75 años de la ciudad A es en promedio superior al del grupo normativo de Tenerife.

Desarrollemos los pasos a seguir en todo contraste de hipótesis.

1) Variable aleatoria

X = Puntaje que obtendrían en el Test Conductual de Memoria de Rivermead (RBMT) las personas entre 70 y 75 años de la ciudad A.

2) Supuestos mínimos necesarios para llevar a cabo el contraste

Los supuestos de un contraste son las afirmaciones que necesitamos establecer (sobre la población de partida y sobre la muestra utilizada) para determinar la distribución de probabilidad en la que se basará nuestra decisión sobre H_0.

Se trata de un contraste unilateral derecho con desvío conocido. En este caso, debido a que el tamaño de la muestra es grande (81 sujetos), no es necesario suponer la normalidad de los puntajes. Por el Teorema Central del Límite, la distribución del estadístico de contraste es aproximadamente normal.

3) Hipótesis estadísticas

El investigador se pregunta si el rendimiento en el test de las personas entre 70 y 75 años de la ciudad A es en promedio *superior* al del grupo normativo de Tenerife.

Debemos plantear las hipótesis estadísticas correspondientes al problema. En realidad, sabemos, todo contraste se basa en la formulación de dos hipótesis: la *hipótesis nula*, simbolizada por H_0, y la *hipótesis alternativa*, simbolizada por H_1.

La hipótesis nula es la que es aceptada provisionalmente como verdadera y que se somete a contraste, es una hipótesis *exacta* (tal cosa es *igual* a tal otra). La hipótesis alternativa, es la negación de la nula, es una hipótesis *inexacta* (tal cosa es *distinta*, *mayor* o *menor* que tal otra).

En nuestro ejercicio, ambas hipótesis se expresan de la siguiente manera:

H_0: $\mu \leq 16{,}1$ El rendimiento en el test RBMT de las personas entre 70 y 75 años de la ciudad A no es superior en promedio al rendimiento en dicho test del grupo normativo de Tenerife.

H_1: $\mu > 16{,}1$ El rendimiento en el test RBMT de las personas entre 70 y 75 años de la ciudad A es superior en promedio al rendimiento en dicho test del grupo normativo de Tenerife.

El procedimiento es el mismo que si las hipótesis fueran:

H_0: $\mu = 16{,}1$
H_1: $\mu > 16{,}1$

<u>Importante</u>: El signo **"="**, tanto si va solo ($\mu = 16{,}1$) como si va acompañado ($\mu \leq 16{,}1$), ***siempre va en la hipótesis nula***. La hipótesis nula, como dijimos, es la que se somete a contraste y es a partir de la afirmación concreta establecida en ella desde donde se inicia todo el proceso de contrastación.

4) Nivel de significación

Necesitamos una *regla de decisión* para decidir si la hipótesis nula planteada debe ser rechazada o no, la misma debe establecerse en términos de *probabilidad*. El criterio que vamos a utilizar con este fin divide a la distribución muestral del estadístico de contraste en dos zonas mutuamente exclusivas: *zona de rechazo* y *zona de aceptación*. La *zona de rechazo* es el área de la distribución muestral que corresponde a los valores del estadístico de contraste que se encuentran tan alejados de lo establecido en H_0, que es muy poco probable que ocurran si H_0, como se supone, es verdadera. La probabilidad de que ocurra esto es **$\propto$,** valor que recibe el nombre de *nivel de significación*. La *zona de aceptación* es el área que corresponde a los valores del estadístico de contraste que es probable que ocurran si H_0, como se supone, es verdadera.

En el ejercicio, trabajamos con un nivel de significación del 5%, es decir, se define una probabilidad de 0,05 como suficientemente improbable.

$$\propto = 0{,}05$$

5) Estadístico de contraste y su distribución bajo H_0

El estadístico de contraste es un resultado muestral que nos brinda información empírica relevante sobre la afirmación propuesta en la hipótesis nula y posee una distribución muestral conocida. En este caso debemos realizar una inferencia sobre **μ**, lo razonable es, entonces, utilizar la información muestral proporcionada por $\overline{X}$.

Sabemos por el Teorema Central del Límite (ya que la muestra es grande) que:

$$E = \frac{\overline{X} - 16{,}1}{\frac{4{,}2}{\sqrt{81}}} \approx N(0{,}1) \text{ , bajo } H_0$$

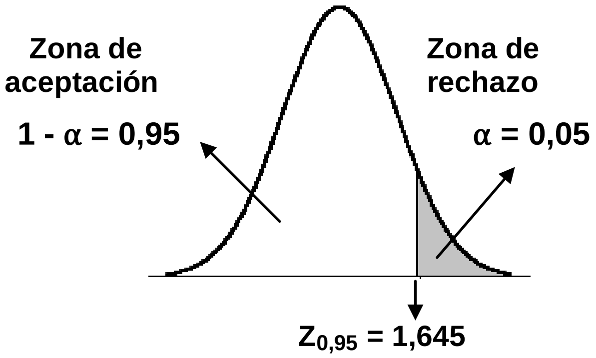
6) Determinar la zona de rechazo de H_0 o “regla de decisión”

La *regla de decisión* consiste en *rechazar* H_0 si el estadístico de contraste toma un valor perteneciente a la *zona de rechazo*; *mantener* H_0 si toma un valor perteneciente a la *zona de aceptación*.

Se trata de un contraste unilateral derecho con desvío conocido. En los contrastes unilaterales, el investigador, o bien posee una idea previa sobre la dirección en la que se producirán los resultados muestrales incompatibles con H_0, o bien considera que sólo son relevantes los resultados muestrales que se muestren incompatibles con H_0 en una de las dos direcciones.

En este caso, el investigador quiere poner a prueba la hipótesis de que el rendimiento en el test de las personas entre 70 y 75 años de la ciudad A es en promedio <u>*superior*</u> al del grupo normativo de Tenerife (contraste unilateral *derecho*); el nivel de significación utilizado es 0,05. Por lo tanto, los valores del estadístico de contraste que se consideran poco probables que ocurran, si H_0 es verdadera, son todos los valores <u>*mayores*</u> que el percentil 95 de la distribución normal estándar.

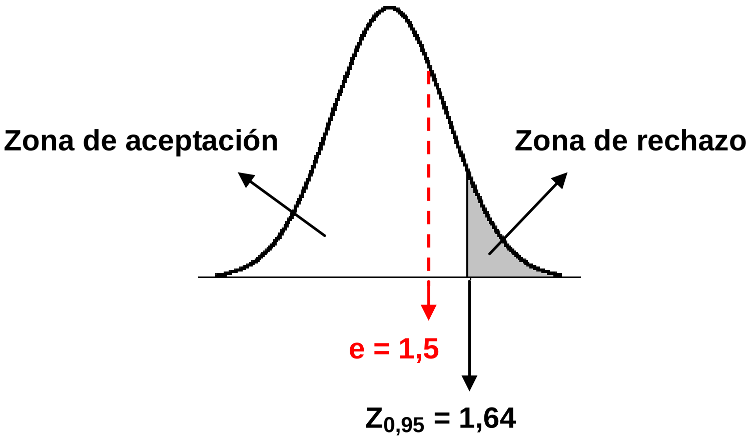
Buscamos en la tabla de la normal, $\mathbf{Z_{0,95} = 1,645}$



Regla de decisión: Se rechaza H_0 si y sólo si el valor observado del estadístico de contraste es mayor que 1,645.

7) Cálculo del valor observado del estadístico de contraste

$$e = \frac{16{,}8 - 16{,}1}{\frac{4{,}2}{\sqrt{81}}} = 1{,}5$$



8) Conclusión

Siendo 1,5 < 1,645, no se rechaza H_0. Los datos observados no dan suficiente evidencia empírica de que el rendimiento en el test de las personas entre 70 y 75 años de la ciudad A sea en promedio superior al del grupo normativo de Tenerife. Si bien la media muestral es mayor que 16,1, esta diferencia no es estadísticamente significativa.

Recordemos que mediante una prueba de hipótesis no se *demuestra* la verdad o falsedad de una hipótesis sino que se decide si los datos observados son consistentes con ella (no la contradicen) o dan suficientes indicios en contra de modo tal que lo más razonable es rechazarla.

b) Determinar qué valores de la media muestral harían la diferencia estadísticamente significativa en la dirección de la hipótesis planteada significa encontrar el punto crítico a partir del cual la diferencia es estadísticamente significativa, en otras palabras, encontrar el valor de la media muestral a partir del cual se rechazaría la hipótesis nula.

En el contraste realizado determinamos que se rechazaría H_0 si y sólo si el valor observado del estadístico de contraste, ***e***, es mayor que 1,645.

Siendo $e = \dfrac{16{,}8 - 16{,}1}{\dfrac{4{,}2}{\sqrt{81}}}$, se rechaza, entonces, H_0 si y sólo si:

$$\frac{\bar{x} - 16{,}1}{\dfrac{4{,}2}{\sqrt{81}}} > 1{,}645$$

Despejando el valor de $\bar{x}$, resulta:

$$\bar{x} - 16{,}1 > 1{,}645 \cdot \frac{4{,}2}{\sqrt{81}} \quad \Rightarrow \quad \bar{x} > 1{,}645 \cdot \frac{4{,}2}{9} + 16{,}1 \quad \Rightarrow \quad \bar{x} > 16{,}867\overset{\frown}{6}$$

$\Longrightarrow$ 16,87 es el valor crítico de la media

Es decir, valores de la media muestral mayores que 16,87 harían la diferencia estadísticamente significativa.

EJERCICIO 2

La impuntualidad de 16 empleados elegidos al azar de cierta dependencia de la administración pública fue de 11 minutos. ¿Son estos datos consistentes con la hipótesis de que la impuntualidad se distribuye normalmente con media 10 y desvío 4 minutos? Usar un nivel de significación del 10%.

Disponemos de los siguientes datos:

	DATOS DE LA POBLACIÓN	DATOS DE LA MUESTRA
Desvío conocido →	μ = 10 σ = 4	$\bar{x} = 11$ n = 16

1) Variable aleatoria

X = Tiempo en minutos en que los empleados de esa dependencia de la administración pública sobrepasa de su horario de entrada.

2) Supuestos mínimos necesarios para llevar a cabo el contraste

- La variable se distribuye normalmente.
- El desvío es conocido ($\sigma = 4$).
- La muestra es aleatoria.

3) Hipótesis estadísticas

H_0: $\mu = 10$ La impuntualidad de los empleados de esa dependencia de la administración pública es en promedio de 10 minutos.

H_1: $\mu \neq 10$ La impuntualidad de los empleados de esa dependencia de la administración pública es en promedio distinta a 10 minutos.

4) Nivel de significación

Nivel de significación del 10%, $\propto = 0{,}10$.

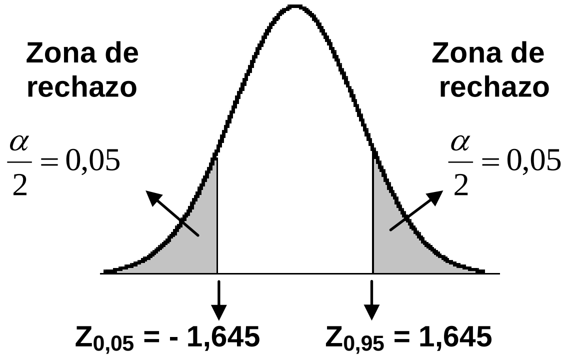
5) Estadístico de contraste y su distribución bajo H_0

$$E = \frac{\overline{X} - 10}{\frac{4}{\sqrt{16}}} \approx N(0{,}1)$$

6) Determinar la zona de rechazo de H_0 o "regla de decisión"

Se trata de un contraste bilateral con desvío conocido. En este caso, la zona de rechazo se encuentra repartida a partes iguales en ambas colas de la distribución muestral. Usamos un valor de significación $\alpha = 0{,}10$, por lo que el área de la zona de rechazo en cada extremo de la distribución muestral es $\frac{\alpha}{2} = 0{,}05$. En consecuencia, hay dos valores críticos que dividen a la distribución en ambas zonas, zona de rechazo y zona de aceptación, ellos son el percentil 5 y el percentil 95.

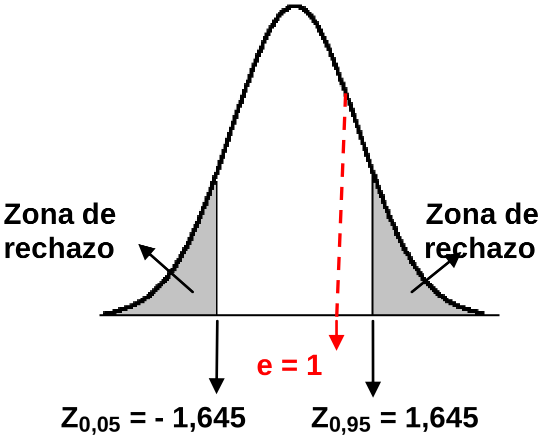
Buscamos en la tabla de la normal, **$Z_{0,05} = -1{,}645$** y **$Z_{0,95} = 1{,}645$**



Regla de decisión: Se rechaza H_0 si y sólo si el valor observado del estadístico de contraste es menor que - 1,645 o mayor que 1,645.

7) Cálculo del valor observado del estadístico de contraste

$$e = \frac{11-10}{\frac{4}{\sqrt{16}}} = 1$$



8) Conclusión

Siendo - 1,645 < 1 < 1,645, no se rechaza H_0. Los datos son consistentes con dicha hipótesis.

EJERCICIO 3

Algunas investigaciones sobre el tiempo de reacción de automovilistas medido como el tiempo en que tardan en frenar desde que perciben un obstáculo inesperado, informan una media de 1,5 segundos con una desviación estándar de 0,2 segundos aproximadamente. Para decidir al 1% si el tiempo de reacción promedio de los porteños es inferior, se tomará una muestra de 64 automovilistas.

a) ¿Para qué valores de la media muestral podría considerarse que el tiempo de reacción promedio de los porteños es inferior al reportado por las investigaciones?
b) Si de la muestra se obtuvo una media de 1,36 segundos, ¿qué concluiría en este caso?
c) Halle el nivel crítico de probabilidad o valor p.

Disponemos de los siguientes datos:

Desvío conocido →

DATOS DE LA POBLACIÓN	DATOS DE LA MUESTRA
μ = 1,5 seg. σ = 0,2 seg.	$\bar{x} = 1{,}36$ n = 64

Se trata de una prueba unilateral a izquierda sobre la media de una variable con desvío conocido, pues se debe decidir al 1% si el tiempo de reacción de los porteños es *inferior* al reportado por las investigaciones (suponiendo que la desviación estándar de la variable para los porteños es aproximadamente la misma que la reportada en la literatura) a)

1) Variable aleatoria

X = Cantidad de segundos que tardan los automovilistas porteños en frenar desde que perciben un obstáculo inesperado.

Nuevamente, la muestra es grande (64 automovilistas), no se necesita suponer la normalidad de los puntajes. Por el Teorema Central del Límite, la distribución del estadístico de contraste es aproximadamente normal.

2) Hipótesis estadísticas

H_0: $\mu = 1,5$ El tiempo de reacción promedio de los porteños no difiere del reportado por las investigaciones.

H_1: $\mu < 1,5$ El tiempo de reacción promedio de los porteños es inferior al reportado por las investigaciones.

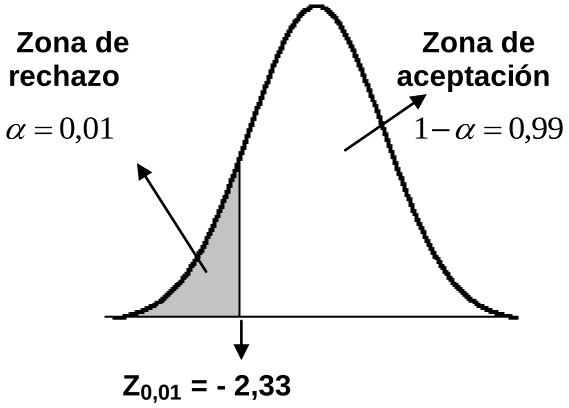
3) Nivel de significación

Nivel de significación del 1%, $\alpha = 0,01$.

4) Estadístico de contraste y su distribución bajo H_0

$$E = \frac{\overline{X} - 1,5}{\frac{0,2}{\sqrt{64}}} \approx N(0,1)$$

5) Determinar la zona de rechazo de H_0 o “regla de decisión”



Regla de decisión: Se rechaza H_0 si y sólo si el valor observado del estadístico de contraste es menor que – 2,33.

Para determinar entonces los valores de la media muestral para los cuáles podría considerarse que el tiempo de reacción promedio de los porteños es inferior al reportado por las investigaciones debemos resolver la siguiente inecuación:

$$\frac{\overline{x} - 1,5}{\frac{0,2}{\sqrt{64}}} < -2,33$$

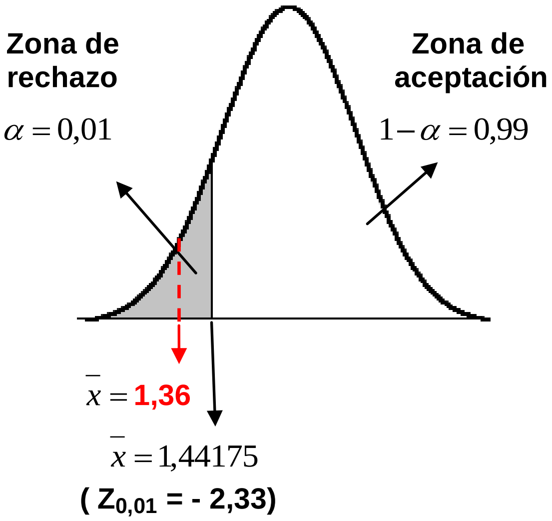
Despejando el valor de $\bar{x}$, resulta:

$$\bar{x} - 1{,}5 < -2{,}33 \cdot \frac{0{,}2}{\sqrt{64}} \quad \Rightarrow \quad \bar{x} < -2{,}33 \cdot \frac{0{,}2}{8} + 1{,}5 \quad \Rightarrow \quad \bar{x} < 1{,}44175$$

Es decir, para valores de la media muestral menores que 1,44175 se podría considerar que el tiempo de reacción promedio de los porteños es inferior al reportado por las investigaciones.

b) Si la media de la muestra es de 1,36 segundos ($\bar{x} = 1{,}36$), este valor es menor que el obtenido en el punto anterior, 1,44175, es decir, si realizamos el contraste de hipótesis correspondiente, con un nivel de significación del 1%, este valor pertenece a la zona de rechazo. Por lo tanto, se concluye que el tiempo medio de los porteños es inferior al reportado.

Representamos gráficamente esta situación:



c) El nivel crítico , lo representamos por *p*, es el nivel de significación más pequeño al que una hipótesis nula puede ser rechazada con el estadístico de contraste obtenido.

Más brevemente, lo definimos como la probabilidad asociada al estadístico de contraste. En este caso, se trata de un contraste unilateral izquierdo, por la tanto, el valor crítico, *p*, será la probabilidad asociada a los valores *menores* que el estadístico de contraste obtenido.

Para hallar el valor de *p* debemos hallar la probabilidad de que el estadístico de contraste sea menor que 1,36, bajo H_0 ($\mu = 1{,}5$).

$$e = \frac{1{,}36 - 1{,}5}{\frac{0{,}2}{\sqrt{64}}} = -5{,}6 \quad \Rightarrow \quad p = P(E < 1{,}36) = P(z < -5{,}6)$$

La tabla normal muestra las áreas bajo la curva para valores de z comprendidos entre - 3 y 3. El área bajo la curva y dichos valores de z es de 0,9974. Como el área total bajo la curva es 1, para valores menores a -3 o mayores a 3, el área es 0,0026. La probabilidad de encontrar un valor tipificado menor a – 3 es, en consecuencia, 0,0013.

Cuanto más se aleja el estadístico de contraste obtenido en esta dirección menor es la probabilidad.

Calculando con Excel la probabilidad de encontrar un valor menor a -5,6 obtenemos $P(z < -5{,}6) = 1{,}07176x10^{-8}$, como vemos es un valor muy cercano a 0.

Si no tenemos la posibilidad de usar este programa decimos, entonces, que ***p* = 0$^+$**

DISTRIBUCIÓN t DE STUDENT

En los ejemplos anteriores hemos visto que, bajo ciertas condiciones, el estadístico $E = \dfrac{\overline{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$ tiene, aproximadamente, una distribución normal estándar.

Ahora bien, el parámetro σ habitualmente no se conoce. La única información que solemos tener sobre la desviación típica poblacional la obtenemos a través de la desviación típica muestral. Si estamos trabajando con muestras grandes, la desviación típica muestral se parecerá a la desviación típica poblacional; de manera que la sustitución de σ por S o S', no hará variar mucho la distribución del estadístico de contraste. Sin embargo, con muestras pequeñas, la sustitución de σ por S o S' tiene consecuencias que no debemos pasar por alto.

En este caso, el estadístico a utilizar, es:

$$E = \frac{\overline{X} - \mu_0}{\dfrac{S'}{\sqrt{n}}}$$

Donde, μ_0 es el valor de μ bajo H_0, S'^2 es la cuasivarianza muestral y S' su raíz cuadrada.

Su distribución, suponiendo que la variable X sea *normal*, sigue otro modelo de probabilidad llamado *t de Student*, con *n* -1 *grados de libertad*, donde *n* es el tamaño de la muestra. Los *grados de libertad* para cualquier estadístico es el número de datos que pueden variar libremente al calcular ese estadístico. Para que E se distribuya según la *t* de *Student*, es necesario, como dijimos, que la población de partida sea *normal* y que la muestra utilizada para obtener $\overline{X}$ sea aleatoria. Dado que nada de esto está afirmado en H_0, necesitamos formularlo como *supuesto*.

Esta distribución es similar a la normal en cuanto a la forma acampanada pero tiene más varianza. A medida que *n* aumenta, S o S' se irán pareciendo más y más a σ, es decir, a medida que *n* aumenta, la distribución de *Student* tiende a la curva normal.

EJERCICIO 4

El gerente de cierta agencia de publicidad desea tomarte a prueba para que dictes un curso de capacitación a sus vendedores asegurándote que sólo cobrarás si el entrenamiento resulta eficaz. Él afirma que hasta el momento el volumen de ventas mensual de sus empleados es en promedio aproximadamente $2150. Antes de aceptar el trabajo decides validar esta afirmación interrogando a 9 vendedores elegidos al azar y obtienes un volumen medio mensual de ventas de $2080 con un desvío estándar de $80.

a) ¿Son estos datos consistentes al 5% con lo informado por el gerente? ¿Aceptarías el trabajo? ¿Por qué?
b) Supongamos que aceptas el trabajo y que entrenas a una muestra aleatoria 16 vendedores, los cuales en los meses subsiguientes del entrenamiento logran un volumen medio mensual de ventas de $2190 con un desvío estándar de $60. Usando el mismo nivel de significación, ¿podrás aspirar al cobro de tu trabajo? Explicita los supuestos necesarios para la resolución de a) y de b).

a)

Disponemos de los siguientes datos:

σ *es desconocida*

DATOS DE LA POBLACIÓN	DATOS DE LA MUESTRA
μ = 2150	$\overline{x} = 2080$ n = 9 S’ = 80

1) Variable aleatoria

X = Volumen de ventas mensual de los empleados de la agencia de publicidad.

2) Supuestos mínimos necesarios para llevar a cabo el contraste

- La variable se distribuye normalmente.
- El desvío es desconocido.
- La muestra es aleatoria.

3) Hipótesis estadísticas

H_0: μ = 2150 El volumen de ventas mensual de los empleados de la agencia de publicidad es en promedio $2150.

H_1: μ $\neq 2150$ El volumen de ventas mensual de los empleados de la agencia de publicidad es en promedio distinto a $2150.

4) Nivel de significación

Nivel de significación del 5%, ∝ = 0,05.

5) Estadístico de contraste y su distribución bajo H_0

$$E = \frac{\overline{X} - 2150}{\frac{80}{\sqrt{9}}} \sim t_8 \text{ bajo } H_0 \quad (\text{gl} = 9 - 1 = 8)$$

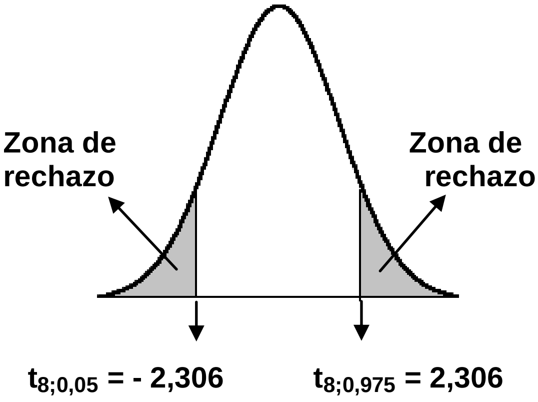
6) Determinar la zona de rechazo de H_0 o “regla de decisión”

Se trata de un contraste bilateral con $\alpha = 0,5$. Los valores del estadístico de contraste que consideramos poco probables que ocurran, si H_0 es verdadera, son todos los valores *menores* que $t_{8;0,025}$ o *mayores* que $t_{8;0,975}$.

La tabla de la Distribución de *Student* muestra sólo los valores de **t** para las probabilidades o áreas mayores a 0,50, pero, siendo una curva simétrica, al igual que en la curva normal, los valores de **t** que dejan a su izquierda o dejan a su derecha un área igual a 0,025 serán iguales en valor absoluto.

Para obtener $t_{8;0,975}$, buscamos la columna correspondiente al nivel de significación 0,975, luego descendemos por esa columna hasta la línea correspondiente a 8 grados de libertad.

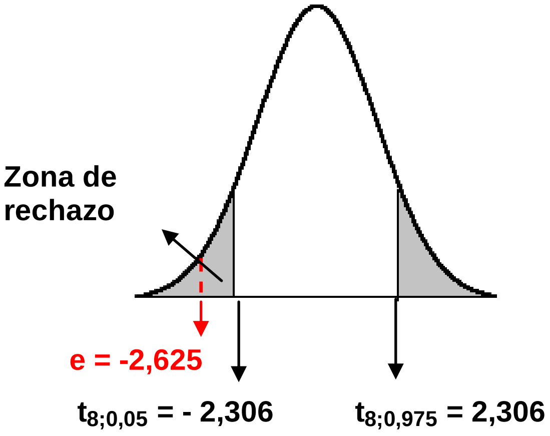
El valor encontrado para t es 2,306. Este valor corresponde a $t_{8;0,975}$. Entonces, $t_{8;0,025} = - t_{8;0,975} = - 2,306$.



Regla de decisión: Se rechaza H_0 si y sólo si el valor observado del estadístico de contraste es <u>menor</u> que – 2,306 o <u>mayor</u> que 2,306.

7) Cálculo del valor observado del estadístico de contraste

$$e = \frac{2080 - 2150}{\frac{80}{\sqrt{9}}} = -2,625$$



8) Conclusión

Siendo - 2,625 < - 2,306, se rechaza H_0. No aceptaría el trabajo porque el gerente está sobreestimando los actuales salarios y eso podría conducir a no detectar si hubiera un verdadero efecto positivo del entrenamiento.

b)

Datos:

	DATOS DE LA POBLACIÓN	DATOS DE LA MUESTRA
σ *es desconocida*	μ = 2150	$\bar{x} = 2190$ n = 16 S’ = 60

1) Variable aleatoria

X = Volumen de ventas mensual de los empleados de la agencia de publicidad.

2) Supuestos mínimos necesarios para llevar a cabo el contraste

- El volumen mensual de ventas que tendrían los vendedores si se entrenaran se distribuye normalmente.
- El desvío es desconocido.
- La muestra es aleatoria.

3) Hipótesis estadísticas

H_0: $\mu = 2150$ El volumen de ventas mensual de los empleados de la agencia de publicidad es en promedio $2150.

H_1: $\mu > 2150$ El volumen de ventas mensual de los empleados de la agencia de publicidad es en promedio mayor a $2150.

4) Nivel de significación

Nivel de significación del 5%, $\propto = 0,05$.

5) Estadístico de contraste y su distribución bajo H_0

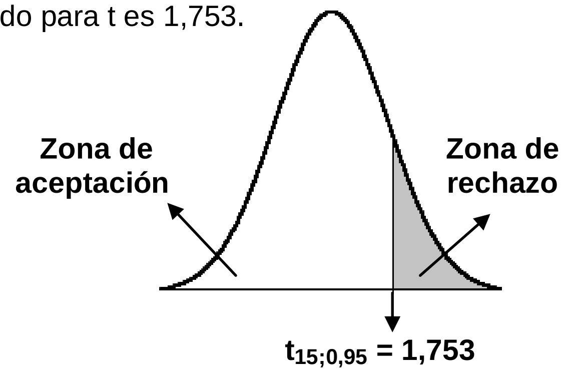
$$E = \frac{\overline{X} - 2150}{\frac{60}{\sqrt{16}}} \sim t_{15} \text{ bajo } H_0 \quad (gl = 16 - 1 = 15)$$

6) Determinar la zona de rechazo de H_0 o “regla de decisión”

Se trata de un contraste unilateral derecho con desvío desconocido y $\alpha = 0,05$, ya que sólo cobrarás el curso de capacitación a los vendedores si el entrenamiento a ellos resulta eficaz, y esto ocurriría si el volumen de ventas de los mismos fuera en promedio superior a $2150. Los valores del estadístico de contraste que consideramos poco probables que ocurran, si H_0 es verdadera, son todos los valores *mayores* que $t_{15;0,95}$.

Para obtener $t_{15;0,95}$, buscamos la columna correspondiente al nivel de significación 0,95, luego descendemos por esa columna hasta la línea correspondiente a 15 grados de libertad.

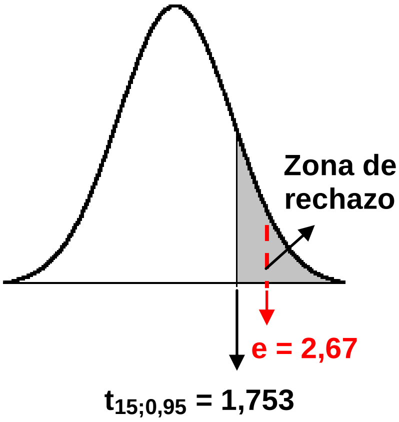
El valor encontrado para t es 1,753.



Regla de decisión: Se rechaza H_0 si y sólo si el valor observado del estadístico de contraste es mayor que 1,753.

7) Cálculo del valor observado del estadístico de contraste

$$e = \frac{2190 - 2150}{\frac{60}{\sqrt{16}}} = 2,67$$



8) Conclusión

Siendo 2,67 > 1,753, se rechaza H_0. Si dictara el curso de capacitación cobraría el trabajo.

EJERCICIO 5

La edad en meses de la menarca de las niñas de cierto distrito del Gran Buenos Aires se distribuye normalmente con una media de 140,5. Un relevamiento reciente sobre las últimas 25 consultas recibidas por el Servicio de Pediatría del hospital zonal dio una edad promedio de 132,2 meses con un cuasi desvío de 21,5 meses. Un equipo interdisciplinario pretende hallar la respuesta a este suceso pero necesitan saber si la disminución en la edad de la menarca registrada es estadísticamente significativa al 10% ¿podrías ayudarlos?

Disponemos de los siguientes datos:

σ *es desconocida*

DATOS DE LA POBLACIÓN	DATOS DE LA MUESTRA
μ = 140,5	$\bar{x} = 132,2$ n = 25 S' = 21,5

1) Variable aleatoria

X = Edad en meses de la menarca de las niñas del Servicio de Pediatría del hospital zonal.

2) Supuestos mínimos necesarios para llevar a cabo el contraste

- La variable se distribuye normalmente.
- El desvío es desconocido.
- La muestra es aleatoria.

3) Hipótesis estadísticas

H_0: $\mu = 140,5$ La edad promedio de la menarca de las niñas del Servicio de Pediatría del Hospital zonal no difiere del de las niñas del distrito del GBA.

H_1: $\mu < 140,5$ La edad promedio de la menarca de las niñas del Servicio de Pediatría del Hospital zonal es inferior al de las niñas del distrito del GBA.

4) Nivel de significación

Nivel de significación del 10%, $\propto = 0,10$.

5) Estadístico de contraste y su distribución bajo H_0

$$E = \frac{\overline{X} - 140,5}{\frac{21,5}{\sqrt{25}}} \sim t_{24} \text{ bajo } H_0 \quad (gl = 25 - 1 = 24)$$

6) Determinar la zona de rechazo de H_0 o "regla de decisión"

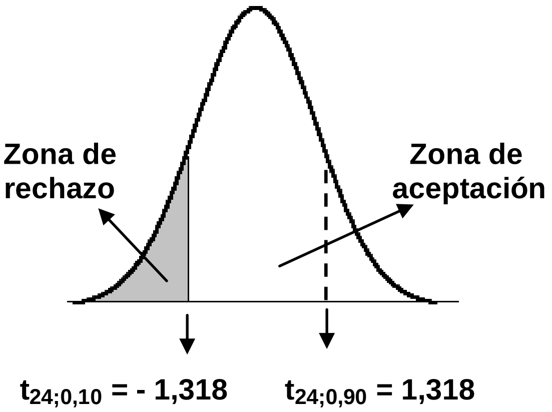
Se trata de un contraste unilateral izquierdo con $\alpha = 0,10$. Los valores del estadístico de contraste que consideramos poco probables que ocurran, si H_0 es verdadera, son todos los valores <u>*menores*</u> que $t_{24;0,10}$.

Buscamos en la tabla de la Distribución de *Student*: $\mathbf{t_{24;0,10}}$.

La tabla muestra sólo los valores de t para las probabilidades o áreas mayores a 0,50, pero, siendo una curva simétrica, al igual que en la curva normal, los valores de t que dejan a su izquierda o dejan a su derecha un área igual a 0,10 serán iguales en valor absoluto.

Para obtener $t_{24;0,10}$, buscamos la columna correspondiente al nivel de significación 0,90, luego descendemos por esa columna hasta la línea correspondiente a 24 grados de libertad.

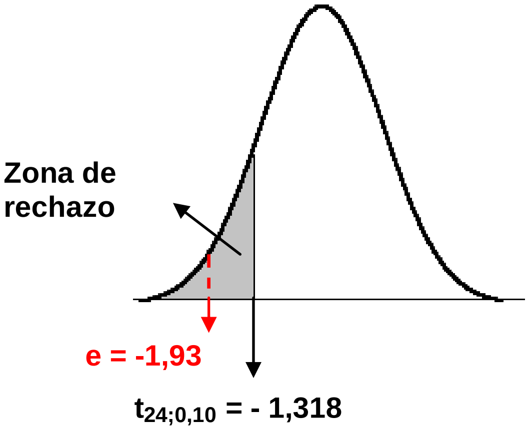
El valor encontrado para t es 1,318. Este valor corresponde a $t_{24;0,90}$. Entonces, $t_{24;0,10} = - t_{24;0,90} = - 1,318$



Regla de decisión: Se rechaza H_0 si y sólo si el valor observado del estadístico de contraste es menor que - 1,318.

7) Cálculo del valor observado del estadístico de contraste

$$e = \frac{132,2 - 140,5}{\frac{21,5}{\sqrt{25}}} = -1,93$$



8) Conclusión

Siendo - 1,93 < - 1, 318, se rechaza H_0. La disminución de la edad media es estadísticamente significativa al 10%.

DIFERENCIA DE MEDIAS DE DOS POBLACIONES RELACIONADAS O APAREADAS

Hablamos de dos muestras relacionadas cuando un grupo de sujetos es evaluado dos veces, por ejemplo, si queremos medir alguna característica en su personalidad antes y después de un determinado estímulo. En estos casos, generalmente, tomamos el valor posterior y le restamos el anterior, para obtener una medida del cambio.

$$D = X_{Después} - X_{Antes}$$

Es importante que esta resta la hagamos de la misma manera con todas las personas de la muestra.

Una vez que tenemos la diferencia de cada persona del estudio, realizamos el resto del procedimiento de prueba de hipótesis como si se tratara de un estudio de una sola muestra de valores, los cuales, en este caso, resultan ser diferencias.

Disponemos, entonces, de una única población (la población de las diferencias), con media μ_D y varianza ${\sigma_D}^2$. Tendremos una única variable *D* con media $\overline{D}$, de la que podremos servirnos para efectuar inferencias sobre μ_D.

De esta forma el estadístico:

$$E = \frac{\overline{D} - \mu_{D_0}}{\dfrac{S'_D}{\sqrt{n}}}$$

se distribuirá, si la población de diferencias es normal o el tamaño muestral lo bastante grande, según el modelo *t* de *Student* con *n* - 1 grados de libertad.
(μ_{D_0} es la media poblacional de dichas diferencias bajo H_0)

Ejemplo:

Un conjunto de sujetos evaluó la culpabilidad del protagonista de un episodio delictivo imaginario en una escala de 1 a 5. Después de proyectar un video con debate sobre las circunstancias inherentes a la vida del personaje volvieron a evaluarlo. A partir de los resultados obtenidos que se muestran a continuación, ¿puede inferir al 5 % que el debate sobre el tema a partir del video modifica la atribución de culpas tendiendo a atenuarla?

	ATRIBUCIÓN DE CULPAS	
Sujeto	Pre-video	Post-video
1	2	2
2	2	1
3	3	2
4	4	3
5	3	3
6	3	2
7	4	4
8	3	2
9	2	1
10	1	2

Para probar si la proyección, con debate, de un video sobre las circunstancias inherentes a la vida del personaje, modifica la atribución de culpas tendiendo a atenuarla, se debe probar estadísticamente al 5 %, si la diferencia entre la atribución de culpas después y antes del debate es en promedio menor a 0.

Buscamos las puntuaciones diferenciales para cada sujeto:

	ATRIBUCIÓN DE CULPAS		
Sujeto	Pre-video $\left(X_A\right)$	Post-video $\left(X_D\right)$	$D_i = X_D - X_A$
1	2	2	0
2	2	1	-1
3	3	2	-1
4	4	3	-1
5	3	3	0
6	3	2	-1
7	4	4	0
8	3	2	-1
9	2	1	-1
10	1	2	1

Desde el punto de vista del análisis de datos, como dijimos anteriormente, disponemos de una única población (la población de las ***diferencias***) con media μ_D y varianza σ^2.

Nos encontramos ante una situación idéntica a la vista para el contraste de hipótesis sobre la media de una población.

1) Variable aleatoria

D = Diferencia de puntajes asignados por los sujetos, después y antes de la proyección del video con debate, respecto de la atribución de culpas al protagonista del episodio delictivo imaginario.

2) Supuestos

- la variable $D = X_D - X_A$ se distribuye normalmente con media μ_D y desvío estándar σ_D desconocido.
- la muestra es aleatoria.

3) Hipótesis estadísticas

$H_0 : \mu_D = 0$ La proyección del video con debate no modifica la atribución de culpas al protagonista del episodio delictivo imaginario.

$H_0 : \mu_D < 0$ La proyección del video con debate atenúa la atribución de culpas al protagonista del episodio delictivo imaginario.

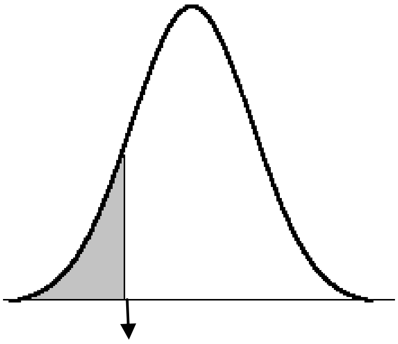
μ_D denota la media de las diferencias entre los puntajes después y antes de la proyección del video con debate, asignados por los sujetos respecto de la atribución de culpas al protagonista del episodio delictivo imaginario.

4) Nivel de significación

Nivel de significación del 5%, $\propto = 0{,}05$.

5) Estadístico de contraste y su distribución bajo H_0

$$E = \frac{\overline{D}}{\frac{S'_D}{\sqrt{10}}} \sim t_9 \text{ bajo } H_0$$



6) Regla de decisión

Es un contraste unilateral izquierdo, se rechaza H_0 si y sólo si el valor observado del estadístico de contraste es menor que $t_{9;0,05}$.

Buscamos en la tabla: $t_{9;0,95} = 1{,}833 \Rightarrow t_{9;0,05} = -1{,}833$

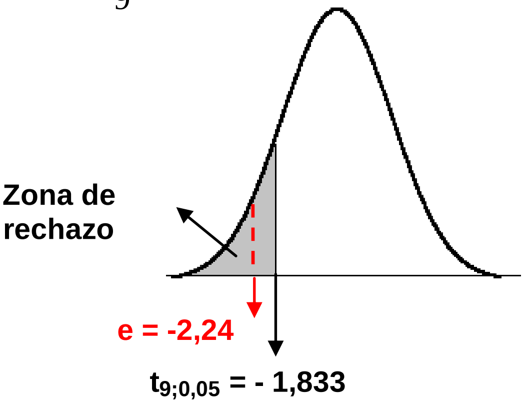
7) Cálculo del valor observado del estadístico de contraste

Calculamos $\overline{D}$ y S'**:**

Sujeto	Pre-video (X_A)	Post-video (X_D)	$D_i = X_D - X_A$	$(D_i - \overline{D})^2$
1	2	2	0	0.25
2	2	1	-1	0.25
3	3	2	-1	0.25
4	4	3	-1	0.25
5	3	3	0	0.25
6	3	2	-1	0.25
7	4	4	0	0.25
8	3	2	-1	0.25
9	2	1	-1	0.25
10	1	2	1	2.25
			-5	4.5

$$\overline{D} = \frac{-5}{10} = -0{,}5 \;, \quad S'^2 = \frac{4{,}5}{9} = 0{,}5 \quad \Rightarrow \quad S' = \sqrt{0{,}5} = 0{,}707$$

$$E = \frac{-0{,}5}{\frac{0{,}707}{\sqrt{10}}} = -2{,}24$$



8) Conclusión

Siendo - 2,24 < - 1,833, se rechaza H_0.

La media de los puntajes en la asignación de culpas es estadísticamente inferior después de la proyección del video con debate que antes, al 5%; lo que podría indicar que los sujetos se vieron en alguna medida influidos por dicha proyección y atenuaron su juicio.

Comentario: Nótese que el nivel de medición es esencialmente ordinal aunque se hace un tratamiento estadístico propio del nivel intervalar. Es uno de los tantos casos donde se comete un cierto abuso en el análisis estadístico el cual podría justificarse si se supone que hay una cierta idea de "equidistancia", aunque no del todo precisada, entre los cinco valores de la escala. Existen métodos estadísticos más adecuados a estas situaciones: por ejemplo, pruebas de hipótesis sobre la mediana, que exceden el nivel de este curso.

CONTRASTE DE HIPÓTESIS SOBRE DIFERENCIA DE MEDIAS INDEPENDIENTES

El contraste de hipótesis sobre dos medias independientes es una técnica de análisis de datos frecuentemente utilizada en la investigación empírica. Si se dan las condiciones apropiadas, el contraste de hipótesis sobre dos medias independientes es el idóneo para comparar dos grupos de sujetos en alguna variable de interés.

Trabajamos con dos poblaciones distintas de las cuales, por lo general, se desconocen las varianzas poblacionales y, por lo tanto, deben estimarse. De dichas poblaciones extraemos, independientemente, dos muestras aleatorias de tamaños n_1 y n_2. Usamos la información muestral para responder si sus respectivas medias son lo bastante diferentes como para pensar que los grupos comparados difieren significativamente en la variable estudiada.

En consecuencia, para efectuar inferencias acerca de μ_1 - μ_2, podemos basarnos en $\overline{X}_1$ - $\overline{X}_2$ y en su distribución muestral y conocer el error de nuestra estimación (igual que ocurría al estimar μ a partir de $\overline{X}$).

Esta prueba funciona de la misma manera que la prueba de hipótesis sobre una media que ya hemos aprendido, con una excepción fundamental: el resultado clave del estudio es una diferencia entre las medias de las dos muestras.

El estadístico a utilizar es:

$$E = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - d}{\sqrt{\frac{(n_1-1)\cdot S_1^{'2} + (n_2-1)\cdot S_2^{'2}}{n_1+n_2-2} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

y se distribuye según el modelo de probabilidad ***t*** de *Student* con $n_1 + n_2 - 2$ grados de libertad.

Supuestos de la prueba *t* para medias independientes:

- Las varianzas poblacionales son desconocidas pero iguales.
- Dos muestras aleatorias simples e independientes.
- Las variables son normales.

EJERCICIO 6 (Datos reales extraídos de una investigación de Ayala et al, 1990).

Para probar la eficacia de un programa de estimulación temprana (PET) se administró un test de C.I. a un grupo control y a un grupo experimental antes de iniciar el programa y después de un año de aplicación del mismo sobre el grupo experimental. Ambos grupos quedaron determinados por asignación aleatoria de niños de 2 a 4 años que asistían al Hogar Infantil “Copetín” de la ciudad de Pamplona, Colombia.

a) Pruebe que no hay diferencias estadísticamente significativas al 10% entre el grupo control y el grupo experimental antes de la implementación del programa.
b) ¿Puede concluirse al 1% que el C.I. aumenta, en promedio, en los niños que reciben el PET? Calcule y valore el tamaño del efecto.
c) No hay evidencias al 1% de que aumente el C.I. medio de los niños que no reciben el PET.
d) ¿Por qué es necesario usar un grupo control? ¿Qué conclusión general puede sacar?

DATOS															
Grupo Control															
Sujeto	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Pre-test	86	135	70	98	111	109	96	92	117	142	107	103	105	116	80
Post-test	110	112	104	106	118	103	110	129	109	117	118	102	124	121	131
Grupo Experimental															
Sujeto	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Pre-test	108	125	105	111	126	105	92	90	104	87	104	104	127	102	100
Post-test	157	147	134	138	154	127	127	129	124	132	124	133	128	144	157

a) Hallamos la media y el cuasidesvío para cada grupo:

<u>Grupo Control</u>:

Sujeto	Pre-test	$\left(x_i - \bar{x}\right)^2$	Sujeto	Pre-test	$\left(x_i - \bar{x}\right)^2$
1	86	341.14	9	117	157.00
2	135	932.08	10	142	1408.50
3	70	1188.18	11	107	6.40
4	98	41.86	12	103	2.16
5	111	42.64	13	105	0.28
6	109	20.52	14	116	132.94
7	96	71.74	15	80	598.78
8	92	155.50			
TOTAL				**1567**	**5099.73**

$$\bar{x} = \frac{1567}{15} = 104{,}47 \quad , \qquad S'^2 = \frac{5099{,}73}{14} = 364{,}27 \quad \Rightarrow \quad S' = \sqrt{364{,}27} = 19{,}0858$$

<u>Grupo Experimental</u>:

Sujeto	Pre-test	$\left(x_i - \overline{x}\right)^2$	Sujeto	Pre-test	$\left(x_i - \overline{x}\right)^2$
1	108	12.46	9	104	0.22
2	125	421.48	10	87	305.20
3	105	0.28	11	104	0.22
4	111	42.64	12	104	0.22
5	126	463.54	13	127	507.60
6	105	0.28	14	102	6.10
7	92	155.50	15	100	19.98
8	90	209.38			
TOTAL				**1590**	**2145.11**

$$\overline{x} = \frac{1590}{15} = 106 \ , \quad S'^2 = \frac{2145{,}11}{14} = 153{,}22 \quad \Rightarrow \quad S' = \sqrt{153{,}22} = 12{,}3782$$

Resumimos los cálculos hechos en el siguiente cuadro:

	$\overline{x}$	S'
Grupo Control	104,47	19,0858
Grupo Experimental	106	12,3782

Debemos probar que no hay diferencias estadísticamente significativas al 10% entre el grupo de control y el grupo experimental antes de la implementación del programa. Se trata de una prueba bilateral sobre una diferencia de medias de variables normales independientes con la misma varianza.

1) Variables aleatorias

X_1: Puntaje en el pre-test de CI de cada niño de la población representada por los niños del grupo control.

X_2: Puntaje en el pre-test de CI de cada niño de la población representada por los niños del grupo experimental.

2) Supuestos

- Los puntajes, X_1 y X_2, se distribuyen normalmente.
- Se extraen dos muestras aleatorias e independientes de las dos poblaciones, cada una de 15 sujetos.
- Las varianzas son desconocidas e iguales.

3) Hipótesis estadísticas

$H_0 : \mu_1 - \mu_2 = 0$ Los sujetos de ambas poblaciones no difieren en promedio en el puntaje C.I. del pre-test.

$H_0 : \mu_1 - \mu_2 \neq 0$ Los sujetos de ambas poblaciones difieren en promedio en el puntaje C.I. del pre-test.

4) Nivel de significación

Nivel de significación del 10%, $\propto = 0{,}10$.

5) Estadístico de contraste y su distribución bajo H$_0$

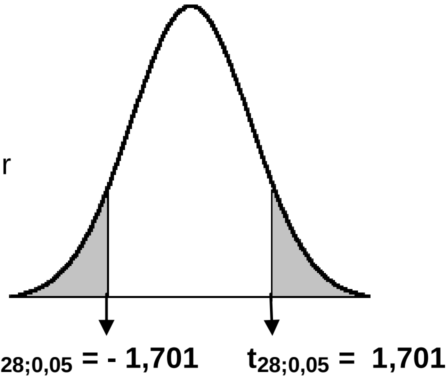
$$E = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{14 \cdot S'^2_1 + 14 \cdot S'^2_2}{15+15-2} \cdot \left(\frac{1}{15} + \frac{1}{15}\right)}} \sim t_{28} \text{ bajo } H_0 \text{ (gl = 15 + 15 – 2 = 28)}$$

6) Regla de decisión

Es un contraste bilateral, se rechaza H_0 si y sólo si el valor observado del estadístico de contraste es menor que $t_{28;0,05}$ o mayor que $t_{28;0,95}$.

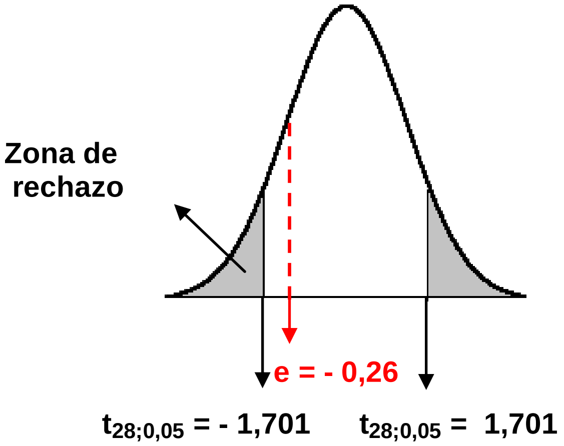
Obtenemos para $t_{28;0,95} = 1{,}701$.

Entonces $- t_{28;0,95} = t_{28;0,05} = - 1{,}701$.



7) Cálculo del valor observado del estadístico de contraste

$$e = \frac{104{,}47 - 106}{\sqrt{\frac{(15-1) \cdot 19{,}0858^2 + (15-1) \cdot 12{,}3782^2}{15+15-2} \cdot \left(\frac{1}{15} + \frac{1}{15}\right)}} = -0{,}26$$



8) Conclusión

Siendo $-1{,}701 < -0{,}26 < 1{,}701$, no se rechaza H_0, por lo que no hay diferencia significativa entre ambos grupos.

b) *Ahora, se debe probar estadísticamente al 1% que, en promedio, los niños aumentarían sus puntajes CI en el post-test si recibieran un programa de estimulación temprana; para lo cual contamos con una muestra de 15 niños sobre los que se llevó a cabo la experiencia (grupo experimental).*

Nuestra hipótesis se refiere, a una población hipotética: la de los niños de 2 a 4 años de características similares a los de la muestra, sometidos a un P.E.T. Se trata de una prueba unilateral a *derecha* para datos pareados.

1) Variable aleatoria

D = Diferencia entre las puntuaciones del post-test y pre-test de cada niño.

2) Supuestos

- la variable D se distribuye normalmente con media μ_D y desvío estándar σ_D desconocido.
- la muestra es aleatoria.

3) Hipótesis estadísticas

$H_0 : \mu_D = 0$ Los CI de los niños después del P.E.T. no difieren, en promedio, de los iniciales.

$H_0 : \mu_D > 0$ Los C.I., en promedio, aumentan si los niños reciben el P.E.T.

μ_D denota la media de las diferencias entre los C.I. obtenidos en el post-test y pre-test.

4) Nivel de significación

Nivel de significación del 1%, $\propto = 0{,}01$.

5) Estadístico de contraste y su distribución bajo H_0

$$E = \frac{\overline{D}}{\frac{S'_D}{\sqrt{15}}} \sim t_{14} \text{ bajo } H_0$$

6) Regla de decisión

Es un contraste unilateral derecho, se rechaza H_0 si y sólo si el valor observado del estadístico de contraste es mayor que $t_{14;0,99}$.

Buscamos en la tabla: $t_{14;0,99} = 2{,}624$

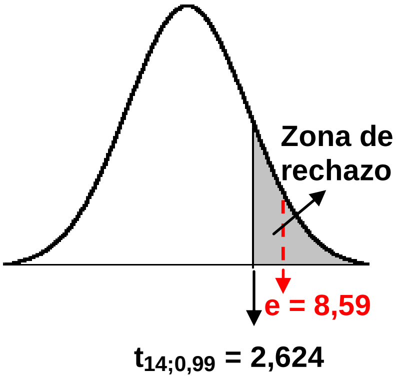
7) Cálculo del valor observado del estadístico de contraste

Sujeto	Pre-test	Post-test	D	$\left(D_i - \overline{D}\right)^2$
1	108	157	49	324
2	125	147	22	81
3	105	134	29	4
4	111	138	27	16
5	126	154	28	9
6	105	127	22	81
7	92	127	35	16
8	90	129	39	64
9	104	124	20	121
10	87	132	45	196
11	104	124	20	121
12	104	133	29	4
13	127	128	1	900
14	102	144	42	121
15	100	157	57	676
TOTAL			**465**	**2734**

Calculamos $\overline{D}$ y S’:

$$\overline{D} = \frac{465}{15} = 31 \;, \quad S'^2 = \frac{2734}{14} = 195{,}2857 \quad \Rightarrow \quad S' = \sqrt{195{,}2857} = 13{,}9745$$

$$E = \frac{31}{\frac{13{,}9745}{\sqrt{15}}} = 8{,}59$$



8) Conclusión

Siendo 8,59 > 2,624, se rechaza H_0. Hubo aumento estadísticamente significativo del CI medio en los niños del grupo experimental

Tamaño del efecto

Una diferencia *estadísticamente significativa* no es necesariamente una diferencia *grande* y tampoco es necesariamente una diferencia *importante*. Un valor estadísticamente significativo de *t* nos permite afirmar que la diferencia entre las medias de las poblaciones representadas por esas dos muestras *no es cero* pero no hay relación entre el valor de *t* y la *magnitud* de la diferencia, porque el valor de *t* no depende solamente de la diferencia entre las dos medias, sino de las varianzas de las muestras y sobre todo del tamaño de las muestras.

Para obviar estos problemas, o para al menos minimizarlos e interpretar mejor los resultados, una de las nuevas técnicas que se van imponiendo es calcular la *magnitud* o *tamaño del efecto*.

Lo que se pretende cuantificar con la *magnitud del efecto* es *en qué grado o en qué medida la hipótesis nula es falsa*. En términos más simples, la *magnitud del efecto* nos permite apreciar si la diferencia *es grande o es pequeña.*

Artículo completo en:
http://www.upcomillas.es/personal/peter/investigacion/Tama%F1oDelEfecto.pdf

d DE COHEN PARA LOS CONTRASTES DE HIPÓTESIS SOBRE MEDIAS

<table>
<tr><th colspan="2">H$_0$: $\mu = \mu_0$</th></tr>
<tr><td>Con desvío conocido σ</td><td>Con desvío estimado s'</td></tr>
<tr><td>$d = \frac{x - \mu_0}{\sigma}$</td><td>$d = \frac{x - \mu_0}{s^{'}}$</td></tr>
</table>

Valoración del tamaño del efecto

Conviene valorar el tamaño del efecto en términos relativos a la información acopiada en torno al problema estudiado. A falta de tal información, Cohen (1988) sugiere el siguiente criterio orientativo: Valores de d entre 0,2 y 0,3 indican un efecto pequeño, alrededor de 0,5 un efecto mediano y mayores que 0,8 un efecto alto.

Aplicando la fórmula *d* de Cohen para contrastes de hipótesis con desvío estimado S

$$d = \frac{31}{13{,}97} = 2{,}22$$

El tamaño del efecto es 2,22; muy alto.

c) *Se debe probar que no hay evidencias al 1% de que aumente el C.I. medio de los niños que no reciben el PET (grupo de control). Es una prueba unilateral a derecha para datos pareados.*

1) Variable aleatoria

D = Diferencia entre las puntuaciones del post-test y pre-test de cada niño que no recibe el PET.

2) Supuestos

- la variable D se distribuye normalmente con media μ_D y desvío estándar σ_D desconocido.
- la muestra es aleatoria.

3) Hipótesis estadísticas

$H_0 : \mu_D = 0$ Los CI de los niños que no reciben P.E.T. no difieren, en promedio, de los iniciales.

$H_0 : \mu_D > 0$ Los C.I. de los niños que no reciben PET aumentan, en promedio.

μ_D denota la media de las diferencias entre los C.I. obtenidos en el post-test y pre-test.

4) Nivel de significación

Nivel de significación del 1%, ∝ = 0,01.

5) Estadístico de contraste y su distribución bajo H_0

$$E = \frac{\overline{D}}{\frac{S'_D}{\sqrt{15}}} \sim t_{14} \text{ bajo } H_0$$

6) Regla de decisión

Es un contraste unilateral derecho, se rechaza H_0 si y sólo si el valor observado del estadístico de contraste es mayor que $t_{14;0,99}$.

Buscamos en la tabla: $t_{14;0,99} = 2,624$

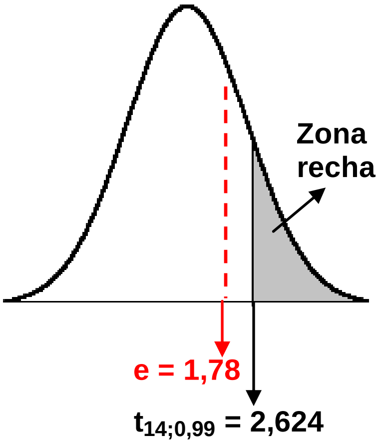
7) Cálculo del valor observado del estadístico de contraste

Sujeto	Pre-test	Post-test	D_i	$(D_i - \overline{D})^2$
1	86	201.64	201.64	201.64
2	135	1075.84	1075.84	1075.84
3	70	585.64	585.64	585.64
4	98	3.24	3.24	3.24
5	111	7.84	7.84	7.84
6	109	249.64	249.64	249.64
7	96	17.64	17.64	17.64
8	92	739.84	739.84	739.84
9	117	316.84	316.84	316.84
10	142	1211.04	1211.04	1211.04
11	107	1.44	1.44	1.44
12	103	116.64	116.64	116.64
13	105	84.64	84.64	84.64
14	116	23.04	23.04	23.04
15	80	1697.44	1697.44	1697.44
Total			**147**	**6332.4**

Calculamos $\overline{D}$ y S':

$$\overline{D} = \frac{147}{15} = 9{,}8 \ , \quad S'^2 = \frac{6332{,}4}{14} = 452{,}3143 \quad \Rightarrow \quad S' = \sqrt{452{,}3143} = 21{,}2677$$

$$E = \frac{9{,}8}{\frac{21{,}2677}{\sqrt{15}}} = 1{,}78$$



8) Conclusión

Siendo 8,59 > 2,624, no se rechaza H_0; por lo que no hay evidencias al 1% de que aumentaría el CI de lo niños que no recibieran el PET.

d) Es necesario usar un grupo control para poder atribuir los cambios observados (en este caso el aumento del CI) al factor que define al grupo experimental: haber recibido el PET. Al haberse probado en a) que ambos grupos estaban en similares condiciones de partida y en b) que aumentaría el CI de quienes recibieran el PET pero en c) que no aumentaría el de quienes no lo recibieran, puede concluirse que el PET es eficaz para aumentar el CI.

Comentario

Para formular la conclusión d), además del contraste de a) que era necesario para probar que ambos grupos partían de condiciones similares, se han llevado a cabo dos contrastes por separado, b) y c), cada uno con su nivel de significación; por lo que la afirmación simultánea de ambas cosas no obedece a un nivel de significación global conocido. Siempre que sea posible, deben tratarse todos los datos del experimento conjuntamente a fin de que la decisión final obedezca a un único nivel de significación. Por otro lado, los niños que no reciban el PET podrían aumentar su CI en alguna medida por la estimulación ordinaria del hogar infantil. De hecho, el lector puede verificar que hay un aumento estadísticamente significativo al 5%. Esto no hace menos cierto que el PET sea eficaz, pues la eficacia del mismo no radica en que los niños que reciban el PET sean los únicos que aumentan su CI sino en que lo aumentan más que si hubieran recibido el programa usual. Teniendo en cuenta esto, sería más acertado plantear que la diferencia entre los CI del post-test y pre-test es, en promedio, mayor para la población hipotética representada por el grupo experimental que para la representada por el grupo control. En símbolos:

$H_0: \mu_{DE} = \mu_{DC}$ En ambas poblaciones la diferencia media de CI del pre al post test es igual.

$H_1: \mu_{DE} > \mu_{DC}$ El aumento, en promedio, del CI de quienes reciben el PET es mayor que el de quienes no lo reciben.

o, equivalentemente,

$H_0: \mu_{DE} – \mu_{DC} = 0$
$H_1: \mu_{DE} – \mu_{DC} > 0$

donde μ_{DE} es la media de la diferencia para la población representada por el grupo experimental y μ_{DC} la correspondiente para el grupo control.

EJERCICIO 7

Los Trastornos de la Conducta Alimentaria (TCA) han despertado un creciente interés durante los últimos años entre los profesionales de la salud debido al incremento de su incidencia y prevalencia[2]. Los TCA tienen una etiología policausal, interactuando de una forma compleja factores predisponentes, factores desencadenantes y mantenedores (Garner y Garfinkel, 1980; Toro y Vilardell, 1987). Dentro de estos factores se ha postulado la insatisfacción con la propia figura (IC) como uno de los factores claves y se encontraron importantes relaciones entre IC y la instauración de TCA.

Un grupo de investigadores realiza un estudio seleccionando dos muestras aleatorias e independientes de 29 chicas y 24 chicos de alrededor de 15 años que cursan el 3º año en una escuela privada de cierto distrito. Les administran el Eating Disorders Inventory (EDI), test original de Garner, Olmsted y Polivy (1983), es un autoinforme que evalúa las características conductuales y cognitivas de la anorexia y la bulimia nerviosa. Consta de 64 ítems agrupados en 8 subescalas. Hay una 2º versión (EDI-2) con 3 subescalas más quedando con 91 ítems. Una de las subescalas mide IC. En esta escala IC los investigadores obtuvieron los siguientes puntajes:

Chicas	13	11	13	10	16	13	13	5	17	17	21	9	8	19	16	2	11	15	10	25	13	22	13	15
	9	15	6	15	10																			
Chicos	8	10	7	3	0	8	16	7	3	12	6	11	4	5	2	4	10	4	10	8	7	8	13	15

¿Es la diferencia observada en la IC entre chicas y chicos estadísticamente significativa al 5%? ¿Es el sexo una FSV? Explicite los supuestos que necesitó hacer para llevar a cabo el contraste. ¿De ser así, cuál es el tamaño del efecto del sexo sobre la IC?

Nota. Este ejercicio, con datos ficticios, está basado sobre la investigación de Baile et al (2003). Artículo completo en http://www.um.es/analesps/v19/v19_2/02-19_2.pdf

Realizamos la tabla de distribución de frecuencias y el cálculo de los estadísticos correspondientes a cada grupo:

---

[2] Términos utilizados en Epidemiología: *Prevalencia* se denomina a la proporción de individuos de un grupo o una población que presentan una característica o evento determinado en un momento o en un período determinado. Refiere a todos los individuos afectados. *Incidencia* es una medida del número de casos nuevos de una enfermedad en un período de tiempo determinado.

Grupo de chicas:

Puntaje EDI	Frecuencia	$x_i \cdot n_i$	$\left(x_i - \overline{x}\right)^2 \cdot n_i$
2	1	2	124.813584
5	1	5	66.781584
6	1	6	51.437584
8	1	8	26.749584
9	2	18	34.811168
10	3	30	30.184752
11	2	22	9.435168
13	6	78	0.177504
15	4	60	13.366336
16	2	32	15.995168
17	2	34	29.307168
19	1	19	33.965584
21	1	21	61.277584
22	1	22	77.933584
25	1	25	139.901584
Total	**29**	**382**	**716.137936**

Calculamos $\overline{x}$ y S'**:**

$$\overline{x} = \frac{382}{29} = 13{,}172 \quad , \quad S'^2 = \frac{716{,}1379}{29} = 25{,}5763 \quad \Rightarrow \quad S' = \sqrt{25{,}5763} = 5{,}0573$$

Grupo de chicos:

Puntaje EDI	Frecuencia	$x_i \cdot n_i$	$\left(x_i - \overline{x}\right)^2 \cdot n_i$
0	1	0	56.881764
2	1	2	30.713764
3	2	6	41.259528
4	3	12	37.637292
5	1	5	6.461764
6	1	6	2.377764
7	3	21	0.881292
8	4	32	0.839056
10	3	30	18.125292
11	1	11	11.957764
12	1	12	19.873764
13	1	13	29.789764
15	1	15	55.621764
16	1	16	71.537764
Total	**24**	**181**	**383.958336**

Calculamos $\bar{x}$ y S'**:**

$$\bar{x} = \frac{181}{24} = 7{,}54 \ , \quad S'^2 = \frac{383{,}9583}{23} = 16{,}6938 \quad \Rightarrow \quad S' = \sqrt{16{,}6938} = 4{,}0858$$

Resumimos los cálculos hechos en el siguiente cuadro:

	$\bar{x}$	S'
Chicas	13,172	5,0573
Chicos	7,54	4,0858

Debemos probar al 5% si la diferencia observada en la IC entre chicas y chicos es estadísticamente significativa. Se trata de una prueba bilateral para la diferencia de medias de variables normales independientes con la misma varianza.

1) Variables aleatorias

X_1: Puntaje en el test IC de cada chica de la población representada por las chicas seleccionadas del 3° año de la escuela privada de ese distrito.

X_2: Puntaje en el test IC de cada chico de la población representada por los chicos seleccionados del 3° año de la escuela privada de ese distrito.

2) Supuestos

- Los puntajes, X_1 y X_2, se distribuyen normalmente.
- Se extraen dos muestras aleatorias e independientes de las dos poblaciones, de 29 y 24 sujetos respectivamente.
- Las varianzas son desconocidas e iguales.

3) Hipótesis estadísticas

$H_0 : \mu_1 - \mu_2 = 0$ Los sujetos de ambas poblaciones no difieren en promedio en el puntaje del test IC.

$H_0 : \mu_1 - \mu_2 \neq 0$ Los sujetos de ambas poblaciones difieren en promedio en el puntaje del test IC.

4) Nivel de significación

Nivel de significación del 5%, ∝ = 0,05.

5) Estadístico de contraste y su distribución bajo H_0

$$E = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{28 \cdot S'^2_1 + 23 \cdot S'^2_2}{29 + 24 - 2} \cdot \left(\frac{1}{29} + \frac{1}{24} \right)}} \quad \sim \quad t_{51} \text{ bajo } H_0 \text{ (gl = 29 + 24 – 2 = 51)}$$

6) Regla de decisión

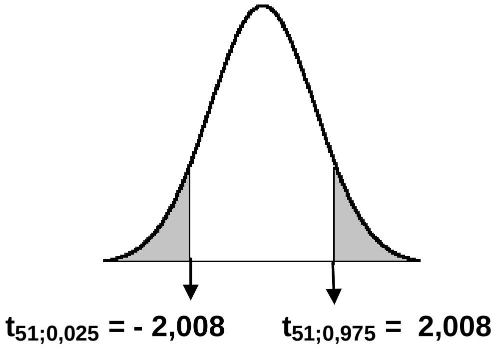
Es un contraste bilateral, se rechaza H_0 si y sólo si el valor observado del estadístico de contraste es menor que $t_{51;0,025}$ o mayor que $t_{51;0,975}$.

Buscando en la tabla $t_{51;0,975}$, vemos que el renglón correspondiente a 51 grados de libertad no está. En la tabla *t* existe una línea para cada grado de libertad, desde 1 hasta el 30. Luego, para cada 10 grados de libertad (30, 40, 50, etc.) hasta 100. Si el estudio incluyera utilizar grados de libertad que se encuentran entre dos valores, como en este caso, utilizamos los grados de libertad más cercanos a los de nuestro problema que se encuentren en la tabla. En el ejercicio deberíamos buscar en el renglón correspondiente a 50 grados de libertad, el valor de **t para 50 gl es 2,009**.

Si buscamos dicho valor t con el programa Excel, obtenemos para $t_{51;0,975}$ = 2,00758373 (redondeando a 3 cifras decimales, es **2,008**).

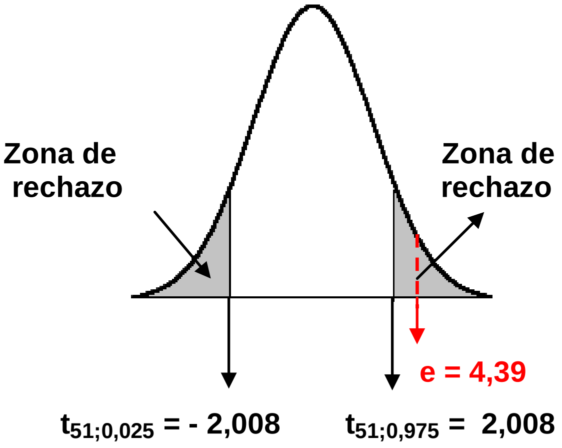
$t_{51;0,975}$ = 2,008.

Entonces - $t_{51;0,975}$ = $t_{51;0,025}$ = - 2,008.



7) Cálculo del valor observado del estadístico de contraste

$$e = \frac{13,172 - 7,54}{\sqrt{\frac{(29-1)\cdot 5,0573^2 + (24-1)\cdot 4,0858^2}{29+24-2}\cdot\left(\frac{1}{29}+\frac{1}{24}\right)}} = 4,39$$



8) Conclusión

Siendo 4,39 > 2,008, se rechaza H_0. El sexo es una fuente de variación, ya que las mujeres, en promedio, presentan mayor insatisfacción con la propia figura.

Tamaño del efecto

La fórmula del *tamaño del efecto g* de Hedges para contrastes de hipótesis sobre diferencia de medias, con desviaciones estándar desconocidas σ$_1$ y σ$_2$, iguales o diferentes, pero estimadas con sus correspondientes S'$_1$ y S'$_2$ es*:*

$$g = \frac{x_1 - x_2}{\sqrt{\dfrac{(n_1 - 1)s_1^{'} + (n_2 - 1)s_2^{'}}{n_1 + n_2 - 2}}}$$

$$g = \frac{13{,}172 - 7{,}54}{\sqrt{\dfrac{(29-1)\cdot 5{,}0573^2 + (24-1)\cdot 4{,}0858^2}{29+24-2}}} = 1{,}21$$

El tamaño del efecto es 1,21; lo cual se considera alto.

EJERCICIO 8

Un investigador desea comprobar si, como afirma la teoría, los sujetos que realizan un dibujo pequeño en el test de la Figura Humana son más introvertidos. El psicólogo posee una muestra de 150 sujetos a los que les administró la escala Introversión Social del Inventario de personalidad MMPI-2. Los puntajes en dicha escala se distribuyen normalmente con media 50 y desviación típica 10. Conformó dos grupos de la siguiente manera: el grupo de introvertidos, compuesto por los sujetos que no superaron los 40 puntos y el grupo de extrovertidos, integrado por los que superaron los 60 puntos. A ambos grupos les administró la técnica de la Figura Humana y midió el tamaño de la representación realizada. La media y el cuasi desvío para cada grupo, expresadas en centímetros, se muestran en la tabla de abajo.

a) Calcule el tamaño de muestra para cada grupo. (Redondee al entero más cercano).

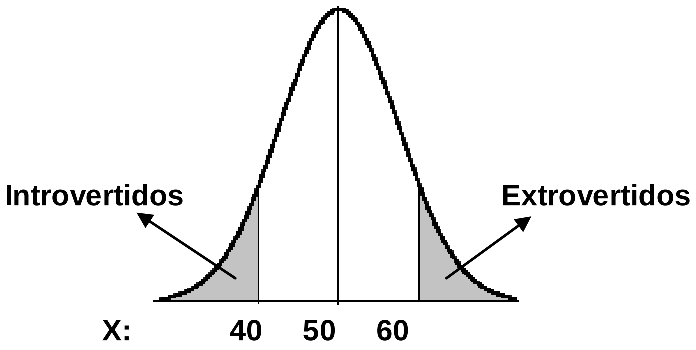
b) ¿Ha podido el investigador comprobar su hipótesis al 5%? Explicite los supuestos necesarios para llevar a cabo el contraste correspondiente. Si su respuesta es afirmativa estime el tamaño del efecto.

	$\overline{X}$	S'
Grupo "Introvertido"	10,3	2,3
Grupo "Extrovertido"	12,2	2,5

a) Sabemos que los puntajes de la escala Introversión Social del Inventario de personalidad MMPI-2 se distribuyen normalmente con media 50 y desviación típica 10 y que los grupos fueron seleccionados de la siguiente manera:

- Introvertidos: sujetos que no superaron los 40 puntos.
- Extrovertidos: sujetos que superaron los 60 puntos.

Gráficamente:



La muestra está conformada por 150 sujetos. Hallamos el área correspondiente a cada grupo, previa tipificación de ambos valores.

Siendo dos puntos simétricos de la curva, ambos valores serán iguales en valor absoluto, pero con signo contrario:

$$Z = \frac{40-50}{10} = -1 \qquad\qquad Z = \frac{60-50}{10} = 1$$

El área a izquierda de Z = -1 es igual al área a derecha de Z = 1.

El área a izquierda de Z = -1 es 0,1587; esto significa que el 15,87% de los sujetos pertenece al grupo de los introvertidos, y también el 15,87% de los sujetos pertenecerá al grupo de los extrovertidos.

Calculamos el 15,87% de 150, esto es aproximadamente 24 sujetos ⇒ $\boldsymbol{n_1 = n_2 = 24}$

b) Se trata de una prueba unilateral izquierda para la diferencia de medias de dos variables normales independientes y con la misma varianza.

1) Variables aleatorias

X_1: Puntaje en el test de la Figura Humana de los sujetos introvertidos.

X_2: Puntaje en el test de la Figura Humana de los sujetos extrovertidos.

2) Supuestos

- Los puntajes, X_1 y X_2, se distribuyen normalmente.
- Se extraen dos muestras aleatorias e independientes de las dos poblaciones, cada una de 24 sujetos.
- Las varianzas son desconocidas e iguales.

3) Hipótesis estadísticas

$H_0 : \mu_1 - \mu_2 = 0$ Los sujetos de ambas poblaciones no difieren en promedio en el puntaje del test de la Figura Humana

$H_0 : \mu_1 - \mu_2 < 0$ Los sujetos introvertidos obtienen en promedio un puntaje inferior en el test de la Figura Humana .

4) Nivel de significación

Nivel de significación del 5%, ∝ = 0,05.

5) Estadístico de contraste y su distribución bajo H_0

$$E = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{23 \cdot S'^2_1 + 23 \cdot S'^2_2}{24+24-2} \cdot \left(\frac{1}{24} + \frac{1}{24}\right)}} \quad \sim \quad t_{46} \text{ bajo } H_0 \ (\text{gl} = 24 + 24 - 2 = 46)$$

6) Regla de decisión

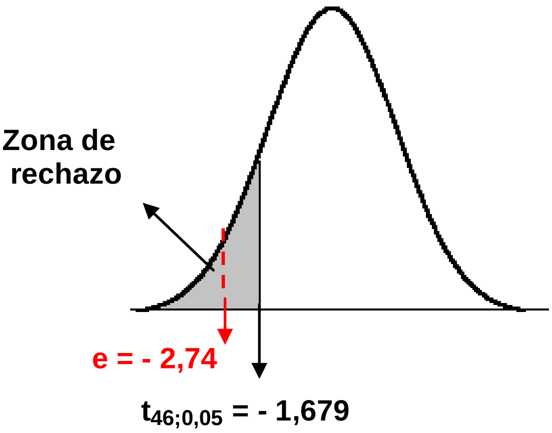
Es un contraste unilateral izquierdo, se rechaza H_0 si y sólo si el valor observado del estadístico de contraste es *menor* que $-t_{46;0,95}$.

Buscando en la tabla $t_{46;0,95}$, vemos que el renglón correspondiente a 46 grados de libertad no está. Utilizamos los grados de libertad más cercanos a los de nuestro problema que se encuentren en la tabla. En el ejercicio deberíamos buscar en el renglón correspondiente a 50 grados de libertad, el valor de t para 50 gl es 1,676

Si buscamos dicho valor t con el programa Excel, obtenemos para $t_{46;0,95}$ = 1,679. Entonces $t_{46;0,05}$ = - 1,679.

7) Cálculo del valor observado del estadístico de contraste

$$e = \frac{10{,}3 - 12{,}2}{\sqrt{\frac{(24-1) \cdot 2{,}3^2 + (24-1) \cdot 2{,}5^2}{24+24-2} \cdot \left(\frac{1}{24} + \frac{1}{24}\right)}} = -2{,}74$$



8) Conclusión

Siendo - 2,74 < - 1,679, se rechaza H_0. El investigador ha podido comprobar su hipótesis.

El tamaño del efecto es:

$$g = \frac{10{,}3 - 12{,}2}{\sqrt{\frac{(24-1)\cdot 2{,}3^2 + (24-1)\cdot 2{,}5^2}{24+24-2}}} = 0{,}79$$

El tamaño del efecto es 0,79, medio alto.

EJERCICIO 9

Efectúe el mismo análisis del Ejercicio Modelo III para el puntaje en satisfacción en las tareas.

Recordemos los datos de dicho artículo:

		Media	Desv. Est.
Satisfacción tareas	Abandonaron	5,42	0,90
	Continúan	5,66	0,80

Debemos responder si los voluntarios que permanecen más de un año están más satisfechos en sus tareas; se trata de una prueba unilateral a derecha para la diferencia de medias de dos variables independientes con varianzas desconocidas pero estimadas con muestras grandes.

1) Variables aleatorias

X_1: Puntaje en el cuestionario de satisfacción de tareas de los voluntarios de organizaciones socioasistenciales que permanecen por lo menos un año en la institución.

X_2: Puntaje en el cuestionario de satisfacción de tareas de los voluntarios de organizaciones socioasistenciales que permanecen menos de un año en la institución.

2) Supuestos

Los puntajes, X_1 y X_2, son independientes. El tamaño de las muestras es grande (502 sujetos permaneció en la institución, 171 abandonaron dentro del transcurso de un año), no es necesario suponer la normalidad de los puntajes.

3) Hipótesis estadísticas

$H_0 : \mu_1 - \mu_2 = 0$ Los sujetos de ambas poblaciones no difieren en promedio en satisfacción en las tareas.

$H_0 : \mu_1 - \mu_2 > 0$ Los sujetos que no abandonan antes del año tienen en promedio más satisfacción en las tareas.

4) Nivel de significación

Nivel de significación del 5%, ∝ = 0,05.

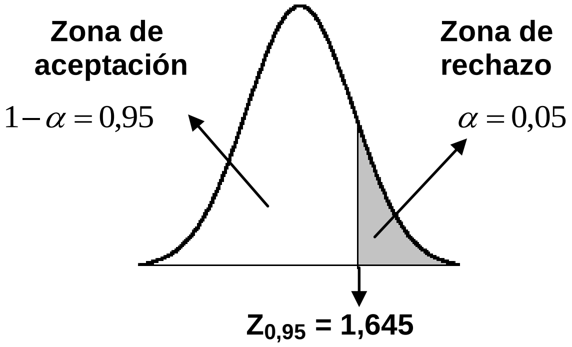
5) Estadístico de contraste y su distribución bajo H_0

$$E = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S'^2_1}{502} + \frac{S'^2_2}{171}}} \sim N(0,1) \quad \text{bajo } H_0$$

Por un teorema similar al Teorema Central del Límite
ya que cada muestra es grande.

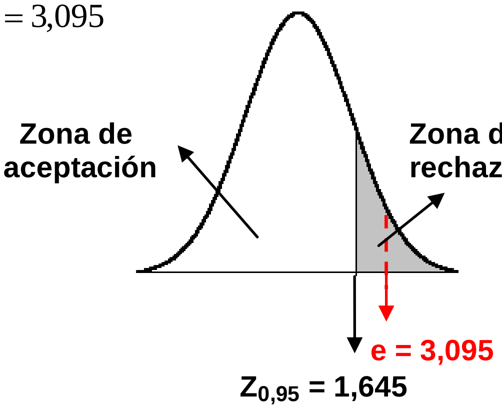
6) Regla de decisión

Es un contraste unilateral a derecha, se rechaza H_0 si y sólo si el valor observado del estadístico de contraste es mayor que $z_{0,95}$.



7) Cálculo del valor observado del estadístico de contraste

$$e = \frac{5,66 - 5,42}{\sqrt{\frac{0,80^2}{502} + \frac{0,90^2}{171}}} = 3,095$$



8) Conclusión

Siendo 3,095 > 1,645, se rechaza H_0.

Tamaño del efecto

$$g = \frac{5,66 - 5,42}{\sqrt{\frac{501 \cdot 0,80^2 + 171 \cdot 0,90^2}{502 + 171 - 2}}} = 0,29$$

El tamaño del efecto es 0,29; es decir pequeño. Los voluntarios que no abandonan muestran en promedio un poco más de satisfacción en las tareas que los que abandonan antes del año.

EJERCICIO 10

Se desea investigar si el nivel de estrés difiere entre las personas que desempeñan tareas ejecutivas y aquellas que trabajan en relación de dependencia. Para ello se tomaron dos muestras aleatorias de 16 ejecutivos y 16 empleados y se les administró un test que mide el nivel de estrés. Por la manera en que se construyó la escala, los psicólogos consideran que para que una diferencia de puntajes en el nivel de estrés sea considerada relevante, ésta debe ser superior a dos puntos. En otras palabras, el umbral es de dos puntos.

Los resultados muestrales dieron un nivel de estrés medio de 55 puntos para los ejecutivos con un cuasi desvío de 6 puntos y una media de 50 con un cuasi desvío de 7 puntos para los empleados. Suponiendo igualdad de varianzas responda:

a) ¿Es la diferencia observada en el nivel de estrés entre ambos grupos estadísticamente significativa al 5%?
b) ¿Puede deducirse de a) la significación psicológica; esto es, que el tipo de actividad tiene un efecto importante sobre el nivel de estrés?
c) ¿Qué hipótesis estadísticas debería plantear para poder concluir que los ejecutivos están significativamente (en sentido psicológico) más estresados que los empleados? Haga la prueba correspondiente.
d) Formule una conclusión general.

Datos del ejercicio:

	N	$\bar{X}$	*S'*
Grupo de ejecutivos	16	55	6
Grupo de empleados	16	50	7

a) Debemos responder al 5% si la diferencia observada en el nivel de estrés es estadísticamente significativa. Se trata de una prueba bilateral.

1) Variables aleatorias

X_1: Nivel de estrés de las personas que trabajan en tareas ejecutivas.

X_2: Nivel de estrés de las personas que trabajan en relación de dependencia.

2) Supuestos

- Los puntajes, X_1 y X_2, se distribuyen normalmente.
- Se extraen dos muestras aleatorias e independientes de las dos poblaciones, cada una de 16 sujetos.
- Las varianzas son desconocidas e iguales.

3) Hipótesis estadísticas

$H_0 : \mu_1 - \mu_2 = 0$ Los sujetos de ambas poblaciones no difieren en promedio en el puntaje del nivel de estrés

$H_0 : \mu_1 - \mu_2 \neq 0$ Los sujetos de ambas poblaciones difieren en promedio en el puntaje del nivel de estrés.

4) Nivel de significación

Nivel de significación del 5%, ∝ = 0,05.

5) Estadístico de contraste y su distribución bajo H$_0$

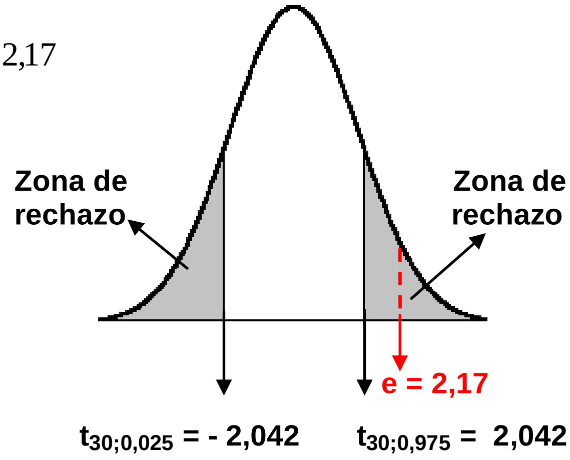
$$E = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{15 \cdot S'^2_1 + 15 \cdot S'^2_2}{16+16-2} \cdot \left(\frac{1}{16} + \frac{1}{16}\right)}} \sim t_{30} \text{ bajo } H_0 \ (gl = 16 + 16 - 2 = 30)$$

6) Regla de decisión

Es un contraste bilateral, se rechaza H$_0$ si y sólo si el valor observado del estadístico de contraste es menor que $t_{30;0,025}$ o mayor que $t_{30;0,975}$.

7) Cálculo del valor observado del estadístico de contraste

$$e = \frac{55-50}{\sqrt{\frac{(16-1)\cdot 6^2 + (16-1)\cdot 7^2}{16+16-2} \cdot \left(\frac{1}{16} + \frac{1}{16}\right)}} = 2{,}17$$



8) Conclusión

Siendo 2,17 > 2,042, se rechaza H$_0$. La diferencia es estadísticamente significativa, mayor para los ejecutivos.

b) No; para que la diferencia tenga una significación psicológica debe superar el umbral, es decir, la diferencia de puntajes en el nivel de estrés debe ser superior a dos puntos.

c)

H_0: $\mu_{EJ} = \mu_{EM} + 2$

H_1: $\mu_{EJ} > \mu_{EM} + 2$ o, equivalentemente,

H_0: $\mu_{EJ} - \mu_{EM} = 2$
H_1: $\mu_{EJ} - \mu_{EM} > 2$

donde μ_{EJ} es La media de estrés en la población de ejecutivos y μ_{EM} la correspondiente en la población de empleados.

Realicemos el contraste:

1) Variables aleatorias

X_1: Puntaje en el nivel de estrés de las personas que trabajan en tareas ejecutivas.

X_2: Puntaje en el nivel de estrés de las personas que trabajan en relación de dependencia.

2) Supuestos

- Los puntajes, X_1 y X_2, se distribuyen normalmente.
- Se extraen dos muestras aleatorias e independientes de las dos poblaciones, cada una de 16 sujetos.
- Las varianzas son desconocidas e iguales.

3) Hipótesis estadísticas

H_0: $\mu_{EJ} - \mu_{EM} = 2$ La diferencia de puntajes en el nivel de estrés en promedio no difiere de 2 puntos entre ambas poblaciones.

H_1: $\mu_{EJ} - \mu_{EM} > 2$ La diferencia de puntajes en el nivel de estrés entre ejecutivos y empleados es en promedio superior a dos puntos.

4) Nivel de significación

Nivel de significación del 5%, $\propto = 0{,}05$.

5) Estadístico de contraste y su distribución bajo H_0

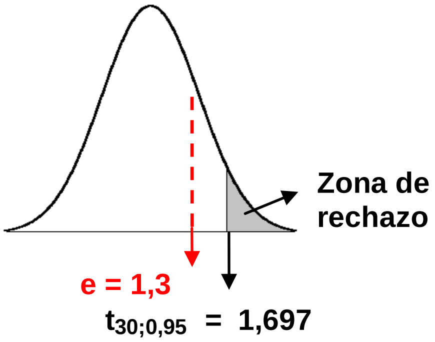
$$E = \frac{\overline{X}_{EJ} - \overline{X}_{EM} - 2}{\sqrt{\frac{15 \cdot S'^2_1 + 15 \cdot S'^2_2}{16+16-2} \cdot \left(\frac{1}{16} + \frac{1}{16}\right)}} \sim t_{30} \quad \text{bajo } H_0 \text{ (gl = 16 + 16 – 2)}$$

6) Regla de decisión

Es un contraste unilateral derecho, se rechaza H_0 si y sólo si el valor observado del estadístico de contraste es mayor que $t_{30;0,95}$.

7) Cálculo del valor observado del estadístico de contraste

$$e=\frac{55-50-2}{\sqrt{\frac{(16-1)\cdot 6^2+(16-1)\cdot 7^2}{16+16-2}\cdot\left(\frac{1}{16}+\frac{1}{16}\right)}}=1{,}3$$



8) Conclusión

Siendo 1,3 < 1,697, no se rechaza H_0. No hay evidencias de que el puntaje medio de estrés en los ejecutivos supere en más de dos puntos al de los empleados.

d) Aunque la diferencia entre los puntajes de ambos grupos es estadísticamente significativa, siendo mayor el puntaje medio de estrés de los ejecutivos; ésta no es psicológicamente significativa porque no hay evidencias de que supere el umbral.

EJERCICIO 11

En el artículo sobre memoria y envejecimiento de Simon et al (2009) se exhibe la siguiente tabla cuyos resultados corresponden a muestras de 22 adultos jóvenes y 22 adultos mayores.

Valoración del agrado suscitado por las imágenes: Media (cuasi desvío)				
Edad	Imágenes			
	Recordadas	No recordadas	Reconocidas	No reconocidas
Jóvenes	3,998 (0,547)	3,991 (0,619)	3,943 (0,518)	3,398 (1,282)
Mayores	5,168 (0,760)	4,982 (0,606)	5,029 (0,613)	4,521 (1,032)
Nota: el agrado se valoró en una escala de 1 (muy desagradable) a 7 (muy agradable)				

Resumiendo la valoración de las imágenes en las cuatro condiciones mediante el promedio se obtiene una media y un cuasi desvío de 3,8325 (0,8053) para los jóvenes y 4,925 (0,7723) para los mayores. Según estos resúmenes ¿puede afirmarse al 1% que la edad es una fuente de variación para la valoración de las imágenes, en qué sentido? De ser así calcule el tamaño del efecto.

Para responder acerca de si la edad es una fuente sistemática de variación debemos realizar un contraste de hipótesis y determinar si existen diferencias estadísticamente significativas respecto de la valoración de las imágenes, en promedio, en mayores y jóvenes.

Es una prueba bilateral para la diferencia de medias de variables normales independientes con igual varianza.

1) Variables aleatorias

X_1: Valoración del agrado suscitado por las imágenes en la población de jóvenes.

X_2: Valoración del agrado suscitado por las imágenes en la población de mayores.

2) Supuestos

- Los puntajes, X_1 y X_2, se distribuyen normalmente.
- Se extraen dos muestras aleatorias e independientes de las dos poblaciones, cada una de 22 sujetos.
- Las varianzas son desconocidas e iguales.

3) Hipótesis estadísticas

$H_0 : \mu_1 - \mu_2 = 0$ La valoración del agrado suscitado por las imágenes no difiere en promedio en ambas poblaciones.

$H_0 : \mu_1 - \mu_2 \neq 0$ La valoración del agrado suscitado por las imágenes difiere en promedio en ambas poblaciones.

4) Nivel de significación

Nivel de significación del 1%, ∝ = 0,01.

5) Estadístico de contraste y su distribución bajo H$_0$

$$E = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{21 \cdot S'^2_1 + 21 \cdot S'^2_2}{22+22-2} \cdot \left(\frac{1}{22} + \frac{1}{22}\right)}} \quad \sim \quad t_{42} \text{ bajo } H_0 \text{ (gl = 22 + 22 – 2)}$$

6) Regla de decisión

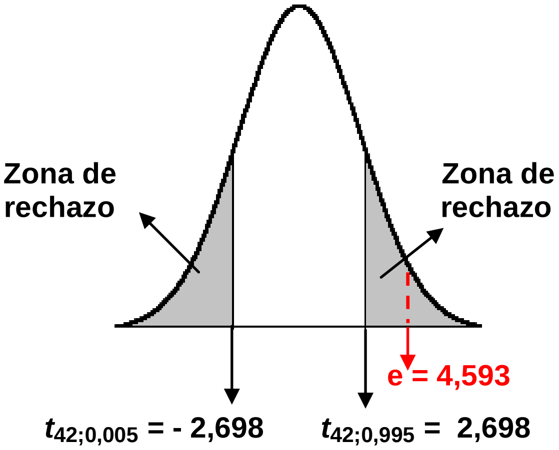
Es un contraste bilateral, se rechaza H$_0$ si y sólo si el valor observado del estadístico de contraste es menor que - $t_{42;0,005}$ o mayor que $t_{42;0,995}$.

Nuevamente nos encontramos con un valor de t para una cantidad de gl que no figuran en la tabla. Buscamos el valor más cercano a 42 gl, esto es el renglón correspondiente a 40 gl, y hallamos: - $t_{42;0,005}$ = - 2,704 ; $t_{42;0,005}$ = 2,704.

Si buscamos dichos valores con Excel, hallamos – 2,698 y 2,698 respectivamente.

7) Cálculo del valor observado del estadístico de contraste

$$e = \frac{4{,}925 - 3{,}8325}{\sqrt{\frac{(22-1) \cdot 0{,}7723^2 + (22-1) \cdot 0{,}8053^2}{22+22-2} \cdot \left(\frac{1}{22} + \frac{1}{22}\right)}} = 4{,}593$$



8) Conclusión

Siendo 4,593 > 2,698, se rechaza H_0; la valoración de las imágenes en los mayores es en promedio más alta; por lo que la edad es una fuente de variación.

Tamaño del efecto

$$g = \frac{4{,}925 - 3{,}8325}{\sqrt{\dfrac{(22-1)\cdot 0{,}7723^2 + (22-1)\cdot 0{,}8053^2}{22+22-2}}} = 1{,}38$$

El tamaño del efecto es 1,38; que es alto.

CONTRASTE DE HIPÓTESIS SOBRE UNA PROPORCIÓN

En el contraste de hipótesis sobre una media nos basamos en la distribución muestral de la media para poder hacer inferencias acerca de la media de una población. Ahora debemos basarnos en la distribución muestral de una proporción.

Consideremos una población cualquiera en la que medimos una variable dicotómica, Se trata de someter a contraste una hipótesis sobre la proporción de éxitos en la población.

Para ello se selecciona una muestra aleatoria de tamaño *n*, y se trata de estudiar si la proporción de éxitos en la muestra se mantiene igual o varió respecto de la proporción de éxitos que se viene dando en la población hasta el momento.

El estadístico a utilizar será ahora:

$$E = \frac{p - \pi_0}{\sqrt{\dfrac{\pi_0 \cdot (1 - \pi_0)}{n}}}$$

donde,

π_0: valor de π bajo la hipótesis nula.

p: proporción observada en la muestra

n: tamaño de la muestra

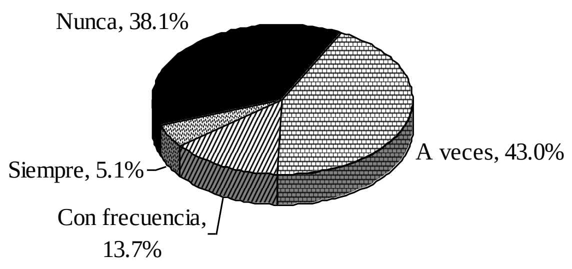
El estadístico E se aproxima a la normal a medida que *n* va aumentando. Podemos, también, por lo tanto, utilizar la distribución normal para diseñar contrastes de hipótesis sobre proporciones.

Los supuestos o condiciones de aplicabilidad son: $n \cdot \pi \geq 5$ y $n \cdot (1-\pi) \geq 5$

EJERCICIO 12

En una investigación sobre conductas indicadoras de posibles problemas de adicción a las nuevas tecnologías, Labrador, F. Villadangos, S. (2010) informan la frecuencia con la que menores escolarizados de la comunidad de Madrid, entre 12 y 17 años, manifiestan tener inclinaciones de dependencia hacia la televisión, Internet, telefonía móvil y video juegos. Para ello tuvieron que responder un cuestionario con escalas Likert de valores Nunca, A veces, Con frecuencia y Siempre. Parte de esos resultados se exhiben a continuación. El artículo completo puede hallarse en http://www.psicothema.com/pdf/3713.pdf

Video Juegos – Sobre el total de menores entre 12 y 17 años



TV	12 años	13 años	14 años	15 años	16 años	17 años
Nunca	90	130	124	174	68	29
A veces	118	120	149	194	56	14
Con frecuencia	50	61	57	68	27	7
Siempre	32	39	40	35	18	10

Se considera que los menores perciben como problemática a una tecnología si sus respuestas son “Con frecuencia” o “Siempre”. A partir de dicha información pruebe las siguientes hipótesis al 1%

a) Menos de la cuarta parte de los menores percibe como problemáticos a los video juegos.
b) Más de la quinta parte de los menores percibe como problemática a la TV.
c) La cuarta parte de los que tienen 14 años percibe como problemática a la TV.

a) La muestra de la investigación está constituida por un total de 1710 menores escolarizados de la comunidad de Madrid, entre 12 y 17 años.

Debemos probar que menos de la cuarta parte de los menores, es decir, menos del 25%, perciben como problemáticos a los videos juegos. Sabemos que los menores perciben como problemática una tecnología si sus respuestas son “Con frecuencia” o

“Siempre”. Según la información que disponemos, un 18,8% (5,1% + 13,7%), dieron estas respuestas, por lo tanto la proporción de éxitos en la muestra es 0,188.

1) Variables y supuestos mínimos necesarios para llevar a cabo el contraste

X: Condición de percibir o no como problemáticos a los video juegos, en la población de menores escolarizados de la comunidad de Madrid, entre 12 y 17 años.

Supuestos: Las observaciones muestrales son independientes una de otra y con la misma probabilidad de percibir como problemáticos a los videos juegos. Es decir, para cada elección al azar, X es una variable Bernoulli con el mismo parámetro π sobre el cual se desea hacer inferencias.

2) Hipótesis estadísticas

$H_0 : \pi \geq 0{,}25$ Por lo menos la cuarta parte de los menores en cuestión percibe a los video juegos como problemáticos.

$H_1 : \pi < 0{,}25$ Menos de la cuarta parte de los menores en cuestión percibe a los video juegos como problemáticos.

El procedimiento es el mismo que si las hipótesis fueran:

$$H_0 : \pi = 0{,}25$$
$$H_1 : \pi < 0{,}25$$

3) Nivel de significación

Nivel de significación del 1%, ∝ = 0,01.

4) Estadístico de contraste y su distribución bajo H_0

$$E = \frac{p - 0{,}25}{\sqrt{\frac{0{,}25 \cdot (1 - 0{,}25)}{1710}}} \approx N(0{,}1) \text{ , bajo } H_0$$

El estadístico E se distribuye normalmente por el Teorema Central del Límite ya que la muestra es grande (1710 niños). Además se cumple que $n \cdot \pi_0 = 1710 \cdot 0{,}25 = 427{,}5 > 5$ y $n \cdot (1 - \pi_0) = 1710 \cdot (1 - 0{,}25) = 1282{,}5 > 5$

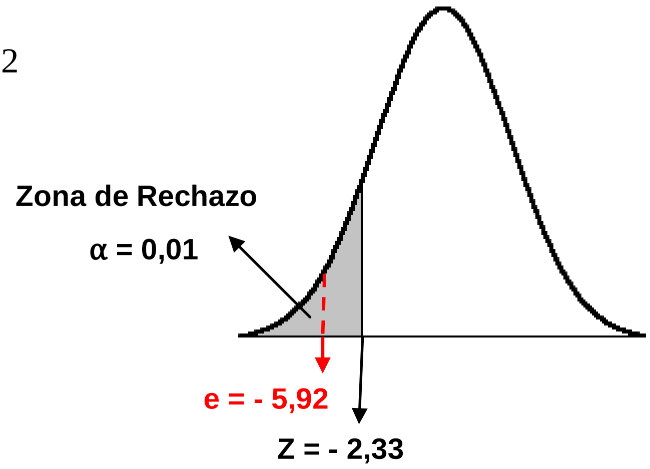
5) Regla de decisión

Es un contraste unilateral izquierdo, se rechaza H_0 si y sólo si el valor observado del estadístico de contraste es menor que $Z_{0,01}$.

Buscamos en la tabla: $Z_{0,01} = - 2{,}33$

6) Cálculo del valor observado del estadístico de contraste

$$e = \frac{0{,}188 - 0{,}25}{\sqrt{\frac{0{,}25 \cdot (1 - 0{,}25)}{1710}}} = -5{,}92$$



7) Conclusión

Como – 5,92 < - 2,33, se rechaza H_0; menos de la cuarta parte percibe como problemáticos los videos juegos.

b) Debemos probar que más de la quinta parte, es decir, más del 20%, percibe como problemática a la TV.

La proporción de éxitos en la muestra, ahora, se obtiene dividiendo el total de niños que dieron como respuesta las opciones “Con frecuencia” o “Siempre” entre el tamaño de la muestra; 270 niños responden “Con frecuencia” y 174, “Siempre” (444 en total).

$$p = \frac{444}{1710} = 0{,}2596$$

Se trata de una prueba unilateral a derecha.

1) Variables y supuestos mínimos necesarios para llevar a cabo el contraste

X: Condición de percibir o no como problemática a la TV, en la población de menores escolarizados de la comunidad de Madrid, entre 12 y 17 años. Es decir, para cada elección al azar, X es una variable Bernoulli con el mismo parámetro π sobre el cual se desea hacer inferencias.

Supuestos: Las observaciones muestrales son independientes una de otra y con la misma probabilidad de percibir como problemática a la TV.

2) Hipótesis estadísticas

$H_0 : \pi \leq 0{,}20$ A lo sumo el 20% de los menores en cuestión percibe a la TV como problemática.

$H_1 : \pi > 0{,}20$ Más del 20% de los menores en cuestión percibe a la TV, como problemática.

El procedimiento es el mismo que si las hipótesis fueran:

$$H_0 : \pi = 0{,}20$$
$$H_1 : \pi > 0{,}20$$

3) Nivel de significación

Nivel de significación del 1%, ∝ = 0,01.

4) Estadístico de contraste y su distribución bajo H_0

$$E = \frac{p - 0{,}20}{\sqrt{\frac{0{,}20 \cdot (1 - 0{,}20)}{1710}}} \approx N(0{,}1) \text{ , bajo } H_0$$

Igual que en el punto a) el estadístico E se distribuye normalmente por el Teorema Central del Límite ya que la muestra es grande, y se cumple que $n \cdot \pi_0 = 1710 \cdot 0{,}20 = 342 > 5$ y $n \cdot (1 - \pi_0) = 1710 \cdot (1 - 0{,}20) = 1368 > 5$

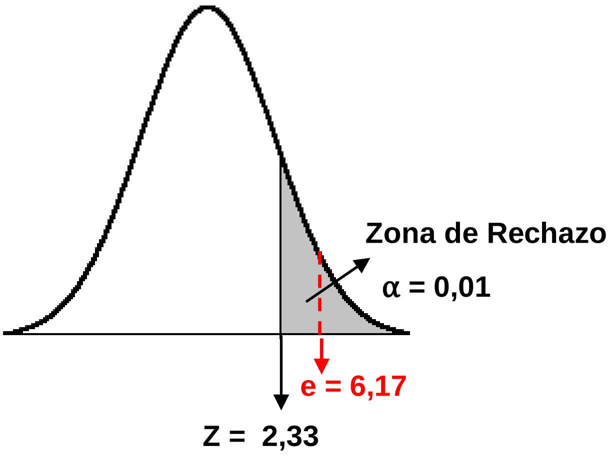
5) Regla de decisión

Es un contraste unilateral derecho, se rechaza H_0 si y sólo si el valor observado del estadístico de contraste es mayor que $Z_{0,99}$.

Buscamos en la tabla: $Z_{0,99} =$ 2,33

6) Cálculo del valor observado del estadístico de contraste

$$e = \frac{0{,}2596 - 0{,}20}{\sqrt{\frac{0{,}20 \cdot (1 - 0{,}20)}{1710}}} = 6{,}17$$



7) Conclusión

Como 6,17 > 2,33, se rechaza H_0; más de la quinta parte percibe como problemática a la TV.

c) Probamos que la cuarta parte, es decir, el 25%, de los que tienen 14 años percibe como problemática a la TV. Se trata de un contraste bilateral. En este caso, la proporción de éxitos de la muestra se obtiene sumando la cantidad de menores, de 14 años, que respondieron “Con frecuencia” o “Siempre” (57 + 40 = 97) y dividiendo este total entre el total de menores, de 14 años (370):

$$p = \frac{97}{370} = 0{,}2621$$

1) Variables y supuestos mínimos necesarios para llevar a cabo el contraste

X: Condición de percibir o no como problemática a la TV, en la población de menores, de 14 años de edad, escolarizados de la comunidad de Madrid. Es decir, para cada elección al azar, X es una variable Bernoulli con el mismo parámetro π sobre el cual se desea hacer inferencias.

Supuestos: Las observaciones muestrales son independientes una de otra y con la misma probabilidad de percibir como problemáticos a los videos juegos.

2) Hipótesis estadísticas

$H_0 : \pi = 0{,}25$ La cuarta parte de los menores en cuestión percibe a la TV como problemática.

$H_0 : \pi \neq 0{,}25$ La cuarta parte de los menores en cuestión no percibe a la TV como problemática.

3) Nivel de significación

Nivel de significación del 1%, ∝ = 0,01.

4) Estadístico de contraste y su distribución bajo H$_0$

$$E = \frac{p - 0{,}25}{\sqrt{\frac{0{,}25 \cdot (1 - 0{,}25)}{1710}}} \approx N(0{,}1) \Big\downarrow$$, bajo H$_0$

Por el Teorema Central del Límite, ya que la muestra es grande, $n \cdot \pi_0 = 370 \cdot 0{,}25 = 92{,}5 > 5$ y $n \cdot (1 - \pi_0) = 370 \cdot (1 - 0{,}25) = 277{,}5 > 5$

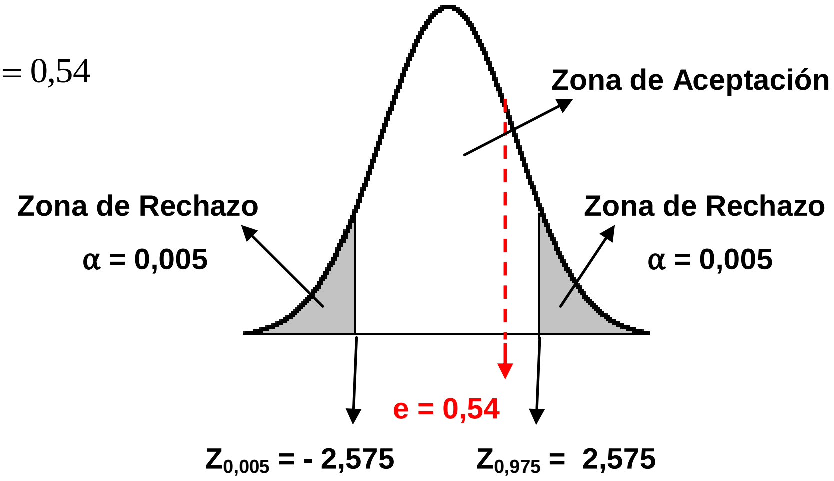
5) Regla de decisión

Es un contraste bilateral, se rechaza H$_0$ si y sólo si el valor observado del estadístico de contraste es menor que Z$_{0,005}$ o mayor que Z$_{0,975}$.

Buscamos en la tabla: Z$_{0,005}$ = - 2,575 y Z$_{0,975}$ = 2,575

6) Cálculo del valor observado del estadístico de contraste

$$e = \frac{0{,}2621 - 0{,}25}{\sqrt{\frac{0{,}25 \cdot (1 - 0{,}25)}{370}}} = 0{,}54$$



7) Conclusión

Como – 2,575 < 0,54 < 2,575, no se rechaza H_0; se valida la afirmación de que la cuarta parte de los jóvenes de 14 años perciben como problemática a la TV.

EJERCICIO 13

Una firma hace un estudio de mercado con el fin de decidir si es necesario o no aumentar la publicidad el siguiente año. Como un aumento de la publicidad implicaría un importante incremento del presupuesto para ese fin, la empresa quiere evitar la posibilidad de lanzar una campaña innecesariamente. La campaña será lanzada sólo si menos del 45% de la población consume el producto de la firma. De 400 personas encuestadas, seleccionadas al azar de distintos distritos de Capital Federal, 168 afirmaron consumir el producto. ¿Qué decisión tomará la firma al nivel del 5%?

Se trata de un contraste unilateral izquierdo.

1) Variables y supuestos mínimos necesarios para llevar a cabo el contraste

X: Condición de consumir o no el producto de la firma, en la población de Capital Federal. Para cada elección al azar, X es una variable Bernoulli con el mismo parámetro π sobre el cual se desea hacer inferencias.

Supuestos: Las observaciones muestrales son independientes una de otra y con la misma probabilidad de consumir o no el producto de la firma.

2) Hipótesis estadísticas

Con el fin de controlar el error de hacer publicidad innecesariamente, deben plantearse las hipótesis:

$H_0 : \pi \geq 0{,}45$ Al menos 45% de las personas consume el producto de la firma.

$H_0 : \pi < 0{,}45$ Menos del 45% de las personas consume el producto de la firma.

El procedimiento es el mismo que si las hipótesis fueran:

$$H_0 : \pi = 0{,}45$$
$$H_0 : \pi < 0{,}45$$

3) Nivel de significación

Nivel de significación del 5%, ∝ = 0,05.

4) Estadístico de contraste y su distribución bajo H$_0$

$$E = \frac{p - 0{,}45}{\sqrt{\frac{0{,}45 \cdot (1 - 0{,}45)}{400}}} \quad \approx \quad N(0{,}1) \text{ , bajo } H_0$$

Por el Teorema Central del Límite, ya que la muestra es grande,
$n \cdot \pi_0 = 400 \cdot 0{,}45 = 180 > 5$ y $n \cdot (1 - \pi_0) = 400 \cdot (1 - 0{,}45) = 220 > 5$

5) Regla de decisión

Es un contraste unilateral izquierdo, se rechaza H$_0$ si y sólo si el valor observado del estadístico de contraste es menor que Z$_{0,05}$.

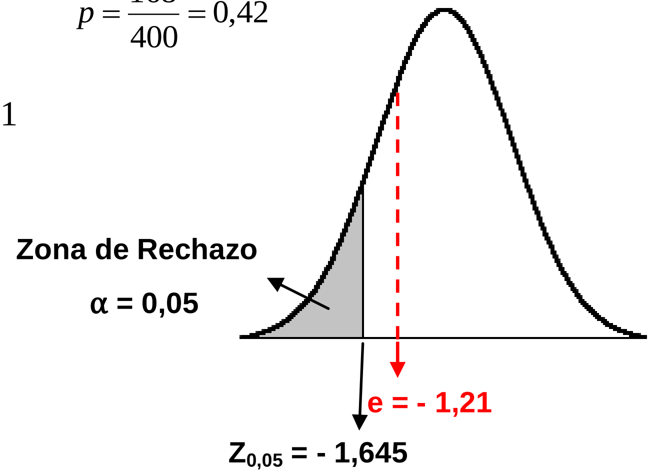
Buscamos en la tabla: Z$_{0,05}$ = - 1,645

6) Cálculo del valor observado del estadístico de contraste

La proporción de éxitos en la muestra se obtiene dividiendo la cantidad de personas encuestadas de la muestra que consume el producto entre el tamaño de la muestra:

$$p = \frac{168}{400} = 0{,}42$$

$$e = \frac{0{,}42 - 0{,}45}{\sqrt{\frac{0{,}45 \cdot (1 - 0{,}45)}{400}}} = -1{,}21$$



7) Conclusión

Como – 1,21 > - 1,645, no se rechaza H$_0$. La empresa decidirá no hacer publicidad por no haber suficiente evidencia de su necesidad.

OBSERVACIÓN: En lo que respecta a medidas de variabilidad, Botella distingue entre los conceptos de varianza y cuasivarianza (la primera se divide el cuadrado de las puntuaciones diferenciales entre *n*, y la segunda entre *n* -1), y desvío estándar y cuasi desvío estándar, para la raíz cuadrada de cada una de ellas, respectivamente. Sin embargo, en investigación, en los artículos científicos, se usa el término “desvío estándar” para ambos (desvío y cuasi desvío), lo mismo que “varianza” para ambas, es decir, cuando se habla de desvío estándar a secas muchas veces se está usando el que divide por n-1 sin hacer la salvedad. Se sabe cuál de los dos es por el contexto donde se está usando. A lo sumo, se habla de “varianza sesgada” (σ_n^2) o “varianza insesgada" (σ_{n-1}^2).

PRÁCTICA DE INTEGRACIÓN
PROBABILIDAD Y ESTADÍSTICA INFERENCIAL

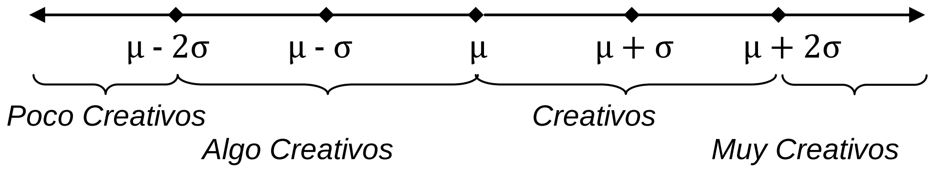
EJERCICIO 1

Los puntajes en un test de creatividad se distribuyen normalmente con media 65 y desvío estándar 10 puntos en un grupo normativo de 600 personas. Los sujetos se consideran según los siguientes criterios: "*Poco creativos*" si exceden dos desvíos por debajo de la media, "*Algo creativos*" si están a menos de dos desvíos por debajo de la media, "*Creativos*" si están a menos de dos desvíos por encima de la media y "*Muy creativos*" si distan más de dos desvíos por encima de la media.

Calcule:

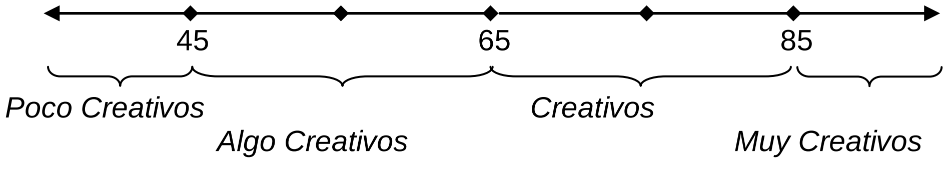
a) la cantidad de individuos del grupo normativo correspondiente a cada categoría.
b) el puntaje mínimo de los individuos que pertenecen al 2% más creativo.
c) cómo es clasificado un sujeto que obtuvo 80 puntos.
d) el puntaje correspondiente al tercer centil.

La situación planteada en el ejercicio es la siguiente:



a) Sabemos que los puntajes en el test de creatividad se distribuyen normalmente con media 65 y desvío estándar 10 puntos. Los sujetos *poco creativos* son los que exceden dos desvíos por debajo de la media, por lo tanto, son los que obtienen un puntaje inferior a 45 (65 - 2·10 = 45); los sujetos *algo creativos* están a menos de dos desvíos por debajo de la media, son los que obtienen un puntaje comprendido entre 45 y 65 puntos; los *creativos*, están a menos de dos desvíos por encima de la media, son los que obtienen un puntaje comprendido entre 65 y 85 puntos (65 + 2·10 = 85); y, finalmente, los *muy creativos*, distan más de dos desvíos por encima de la media, es decir, obtienen un puntaje mayor a 85.

Gráficamente,

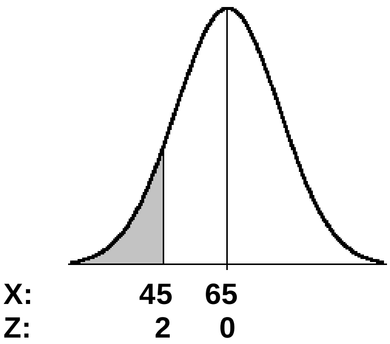


Haciendo uso de la distribución normal, transformamos cada uno de estos puntajes, en puntaje z y, luego, hallamos el área correspondiente a cada categoría.

$$z = \frac{45-65}{10} = -2 \qquad\qquad z = \frac{85-65}{10} = 2$$

La media del test de creatividad, 65, se corresponde con z = 0.

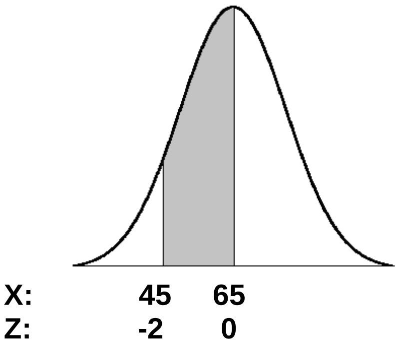
Sujetos Poco Creativos:



Buscamos en la tabla el área a izquierda de z = - 2. Dicha área es 0,0228; el porcentaje de sujetos que se corresponde con el área hallada se obtiene multiplicando este valor por 100 esto significa que el 2,28% (0,0228 · 100) de la población corresponde a la categoría *Poco Creativos*.

El 2,28% de 600 es 13,68 (14 sujetos).

Sujetos Algo Creativos:



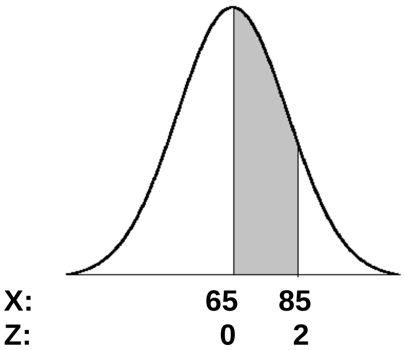
Para calcular el área encerrada por ambos valores, restamos de 0,5 (área a izquierda de z = 0) 0,0228 (área a izquierda de z = - 2).

0,5 – 0,0228 = 0,4772

El porcentaje de sujetos algo creativos es 47,72%.

El 47,72% de 600 es 286,32 (286 sujetos).

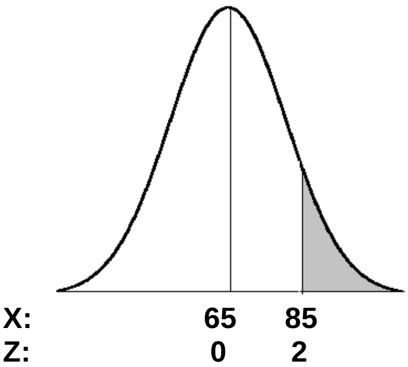
Sujetos Muy Creativos:



Recordemos que la curva normal es simétrica, en consecuencia el área comprendida entre los puntajes 65 y 85 (o entre 0 y 2) es igual al área comprendida entre los puntajes 45 y 65 (o entre -2 y 0), es decir, 0,4772.

La cantidad de sujetos creativos es también 286.

Sujetos Muy Creativos:

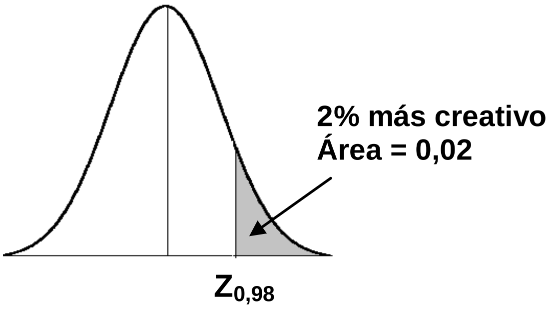


Análogamente, el área a derecha de z = 2 es 0,0228.

La cantidad de sujetos muy creativos es 14.

Poco Creativos	Algo Creativos	Creativos	Muy Creativos
14	286	286	14

b) Debemos buscar el puntaje que corresponde al 2% más creativo. La representación gráfica de esta situación es:



Buscamos en la tabla el valor de z que deja a izquierda un área igual a 0,98 (1 – 0,02). Dicho valor es 2,05. El puntaje en el test de creatividad se obtiene despejando de la fórmula:

$$z = \frac{x-\mu}{\sigma}$$

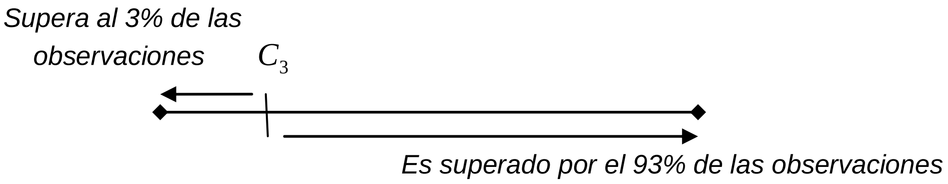
$$2{,}05 = \frac{x-65}{10} \quad \Rightarrow \quad 2{,}05 \cdot 10 + 65 = x \quad \Rightarrow \quad x = 85{,}5$$

El puntaje que corresponde al 2% más creativo es 85,5.

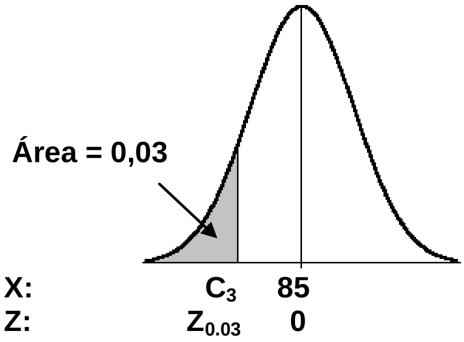
c) *Creativo* pues está a un desvío y medio por encima de la media.

d) Recordemos el concepto de centil. El tercer centil es la puntuación que supera el 3% de las observaciones y es superada por el 97% de las mismas.

Gráficamente:



Los puntajes del test de creatividad se distribuyen normalmente, por lo tanto podemos hacer uso de la curva normal para obtener C$_3$.



$$-1,88 = \frac{x-65}{10} \quad \Rightarrow \quad -1,88\cdot 10+65 = x \quad \Rightarrow \quad x = 46,2$$

C$_3$ = 46,2

EJERCICIO 2

Sabiendo por experiencia que sólo 4 de cada 10 personas que comienzan un determinado tratamiento perseveran en él, ¿cuál es la probabilidad de que en una muestra aleatoria de 8 personas que comienzan el tratamiento

a) persevere más de la mitad?
b) persevere por lo menos uno?
c) no persevere a lo sumo la mitad?

a) La muestra está constituida por 8 personas (***n* = 8**). Para cada persona seleccionada tenemos dos resultados posibles: que persevere en dicho tratamiento o no. Son resultados mutuamente excluyentes. Consideremos como éxito la condición de seleccionar una persona que persevere en el tratamiento. Sabemos que 4 de cada 10 personas perseveran en él, por lo tanto, la probabilidad de éxito es 0,4 (***p* = 0,4**).

Definimos la variable aleatoria X:

X = Cantidad de personas que persevera en el tratamiento.

La variable X se distribuye binomialmente con parámetros *n* = 8 y *p* = 0,4.

X ~ B (8; 0,4)

Hallar la probabilidad de que persevere en el tratamiento más de la mitad significa calcular la probabilidad de que perseveren más de 4 personas:

$$P(X>4) = P(X=5)+P(X=6)+P(X=7)+P(X=8) =$$

= 0,124 + 0,041 + 0,008 + 0,001 = **<u>0,174</u>**

b) Hallar la probabilidad de que persevere por lo menos uno equivale a:

$P(X > 1) = 1 - P(X = 0) = 1 - 0{,}017 = 0{,}983$

<u>0,983</u>

c) Hallar la probabilidad de que no persevere a lo sumo la mitad es equivalente a calcular la probabilidad de que persevere como mínimo la mitad:

$P(X \geq 4) = P(X = 4) + P(X = 5) + P(X = 6) + P(X = 7) + P(X = 8) =$

= 0,232 + 0,124 + 0,041 + 0,008 + 0,001 = **<u>0,406</u>**

a) 0,174 b) 0,983 c) 0,406

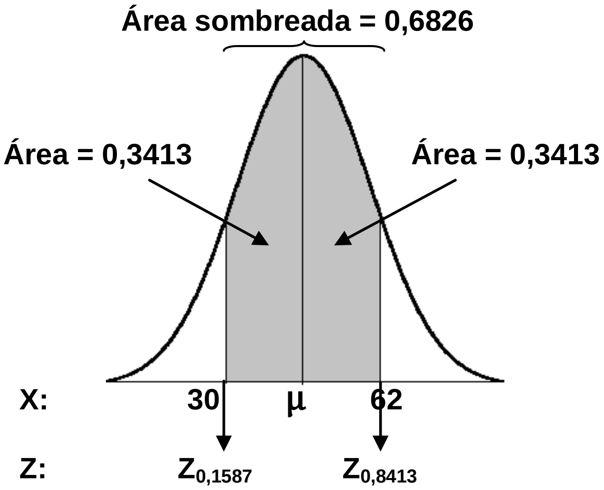
EJERCICIO 3

En una población determinada se incluye en la “clase activa” a todas aquellas personas que superan los 18 años y que no superan los 70 años. Los que han quedado por fuera de este intervalo son incluidos en la categoría “clase pasiva”. Sabiendo que el 68,26% de la población (en torno a la media) oscila entre 30 y 62 años y que la distribución de las edades es normal calcular:

a) La media y el desvío poblacional de la distribución.
b) La probabilidad de que una persona elegida al azar pertenezca a la clase pasiva.
c) El porcentaje de personas con más de 23 años.
d) La edad mínima necesaria para acceder a una asignación especial que el gobierno dispone entregar por única vez al 20% de las personas de más edad.
e) La probabilidad de que en una muestra aleatoria de 10 personas a lo sumo 7 no estén en condiciones de acceder a la asignación especial de d).

a) El 68,26% de la población en torno a la media ocupa un área central, en la distribución normal, igual a 0,6826. La misma se distribuye en partes iguales, 0,3413, a izquierda y a derecha de μ**.** Las edades que limitan esta región son 30 y 62 años.

Gráficamente:



El área total bajo la curva es 1, por lo tanto, el área sin sombrear, en cada cola de la distribución, es igual a:

(1 – 0,6826) : 2 = 0,1587

Los valores de Z que limitan la zona central son:

$$Z_{0,1587} = -1 \qquad Z_{0,8413} = 1$$

Reemplazando los datos en $z = \frac{x-\mu}{\sigma}$, obtenemos un sistema de ecuaciones con dos incógnitas, μ y σ:

$$\begin{cases} 1 = \frac{62-\mu}{\sigma} \\ -1 = \frac{30-\mu}{\sigma} \end{cases}$$

Resolvemos el sistema mediante el método de sustitución.

Despejamos σ en la primera ecuación: $\sigma = 62-\mu$ (1)

Reemplazamos dicha variable en la segunda ecuación:

$$-1 = \frac{30-\mu}{62-\mu} \quad \Rightarrow \quad -1\cdot(62-\mu) = 30-\mu \quad \Rightarrow \quad -62+\mu = 30-\mu \quad \Rightarrow \quad 2\mu = 92$$

μ = 46

Reemplazamos el valor hallado para μ en (1):

σ = 16

b) Pertenecen a la clase pasiva aquéllas personas que tienen 18 años o menos y los que superan los 70 años. Si llamamos X a la variable "Edad de cada persona en esa población", la probabilidad de que una persona pertenezca a la clase pasiva es:

$$P(X \le 18) + P(X > 70)$$

Los puntajes Z correspondiente a cada valor son:

$$z = \frac{18-46}{16} = -1{,}75 \qquad z = \frac{70-46}{16} = 1{,}5$$

Luego,

$$P(X \le 18) = P(Z \le -1{,}75) = 0{,}0401$$

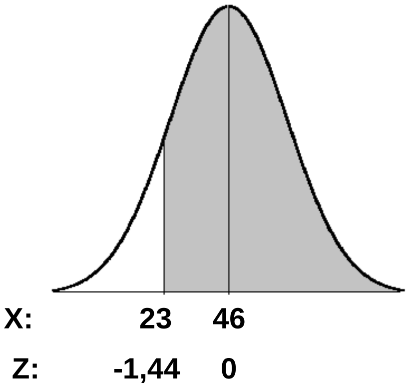
$$P(X > 70) = P(Z > 1{,}5) = 1 - P(Z < 1{,}5) = 1 - 0{,}9332 = 0{,}0668$$

La probabilidad de que una persona pertenezca a la clase pasiva es la suma de ambas probabilidades:

0,0401 + 0,0668 = **0,1069**

c) Hallar el porcentaje de personas con más de 23 años equivale a hallar el área que se encuentra a la derecha de la puntuación 23 y luego multiplicar dicho valor por 100.

Gráficamente:



$$z = \frac{23 - 46}{16} \approx -1{,}44$$

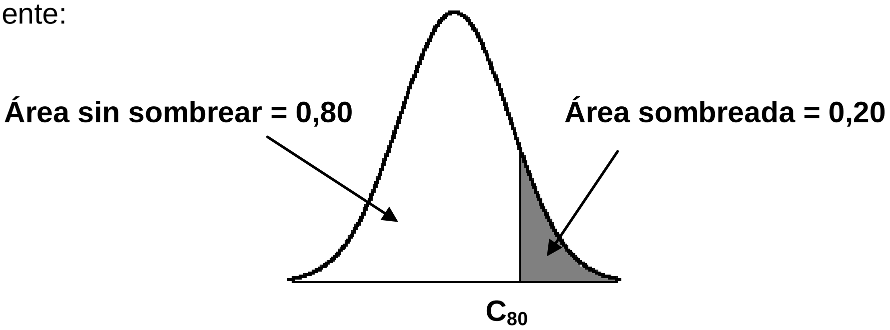
z = – 1,44 deja un área a izquierda de 0,0749, por lo tanto, el área a derecha es:

1 – 0,0749 = 0,9251

Multiplicando este valor por 100, obtenemos el porcentaje de personas que tienen más de 23 años: **92,51%**

d) Debemos encontrar la puntuación que es superada por el 20% de las observaciones y, por lo tanto, supera al 80%. Dicha puntuación es el C_{80}.

Gráficamente:



Buscamos en la tabla de la normal $Z_{0,80}$, es decir, el valor de Z que deja a izquierda un área igual a 0,80.

$$Z_{0,80} = 0{,}84 \quad \Rightarrow \quad 0{,}84 = \frac{x - 46}{16} \quad \Rightarrow \quad 0{,}84 \cdot 16 + 46 = x \quad \Rightarrow \quad x = 59{,}44$$

60 años

Acceden a la asignación especial que ofrece el gobierno las personas con una edad mayor o igual a 59,44, es decir, mayor o igual a 60.

e) Debemos hallar la probabilidad de que en una muestra aleatoria de 10 personas a lo sumo 7 no estén en condiciones de acceder a la asignación especial que ofrece el gobierno. Una persona, para acceder a dicha asignación debe tener una edad mayor o igual a 60 años, según vimos en el punto anterior, y la probabilidad que tiene una persona de acceder a ella es 0,20.

La muestra está compuesta por 10 sujetos, ***n* = 10**. Cada sujeto seleccionado tiene una edad mayor o igual a 60 años o no (tenemos sólo dos resultados posibles). Consideramos éxito la condición de tener una edad mayor o igual a 60 años.

Sea X = “Cantidad de sujetos que tienen una edad mayor o igual a 60”

La variable X se distribuye binomialmente con parámetros *n* = 10 y probabilidad de éxito *p* = 0,20:

X ~ B(10; 0,20)

Hallar la probabilidad de que a lo sumo 7 personas no accedan a la asignación es equivalente a hallar la probabilidad de que sí accedan un mínimo de tres personas, es decir, X debe ser mayor o igual a 3.

$$P(X \geq 3) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)]$$

= 1 – [0,107 + 0,268 + 0,302] = 1 – 0,678 = **0,322**

EJERCICIO 4. Artículo completo en http://www.psicothema.com/pdf/3496.pdf

Orgilés, M., Espada, J. y Méndez, F. (2008) administraron el Cuestionario de Ansiedad por Separación Infantil, Forma Niños, CASI-N (Espada et al, 2006) a 95 niños cuyos padres se encontraban separados o divorciados y a 95 niños que convivían con ambos progenitores por no haberse producido una ruptura de pareja. La media y desviación estándar para los primeros fue 29,66 y 11,49 mientras que para los segundos 26,99 y 10,83. ¿Hay diferencias estadísticamente significativas al 10% entre los niños de ambas condiciones? De ser así estime el tamaño del efecto e indique en qué sentido se da tal diferencia.

Datos del ejercicio:

	n	$\bar{X}$	*S'*
Niños cuyos padres están separados o divorciados	95	29,66	11,49
Niños que conviven con ambos progenitores	95	26,99	10,83

Debemos responder al 10% si la diferencia observada en el nivel de ansiedad por separación infantil es estadísticamente significativa. Se trata de una prueba bilateral.

1) Variables aleatorias

X_1: Puntaje de cada niño en el Cuestionario de Ansiedad por Separación Infantil en los niños cuyos padres están separados o divorciados.

X_2: Puntaje de cada niño en el Cuestionario de Ansiedad por Separación Infantil en los niños que conviven con ambos progenitores.

2) Supuestos

Como las muestras seleccionadas son grandes, no se necesita suponer la normalidad de los puntajes en cada población ni la homogeneidad de las varianzas.

3) Hipótesis estadísticas

$H_0 : \mu_1 - \mu_2 = 0$ Los niños de ambas poblaciones no difieren en promedio en el puntaje del nivel de Ansiedad por Separación Infantil.

$H_0 : \mu_1 - \mu_2 \neq 0$ Los niños de ambas poblaciones difieren en promedio en el puntaje del nivel de Ansiedad por Separación Infantil.

4) Nivel de significación

Nivel de significación del 10%, $\propto = 0,10$.

5) Estadístico de contraste y su distribución bajo H$_0$

$$E = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^2}{95} + \frac{S_2^2}{95}}} \approx N(0,1)$$

por un teorema similar al Teorema Central del Límite ya que cada muestra es grande.

Bajo H$_0$

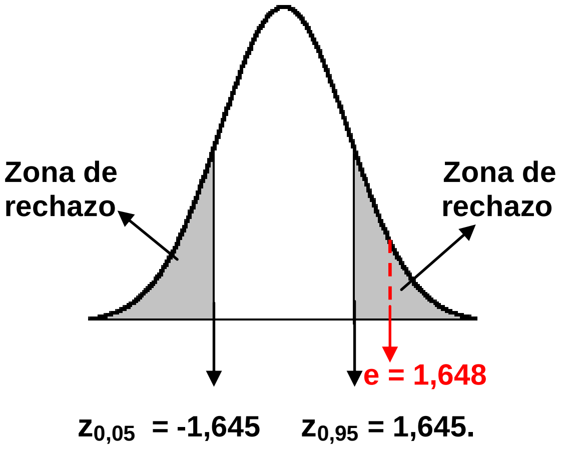
6) Regla de decisión

Es un contraste bilateral, se rechaza H$_0$ si y sólo si el valor observado del estadístico de contraste es menor que $z_{0,05}$ o mayor que $z_{0,95}$.

Buscamos en la tabla de la normal los valores de z: $z_{0,05} = -1,645$, $z_{0,95} = 1,645$.

7) Cálculo del valor observado del estadístico de contraste

$$e = \frac{29,66 - 26,99}{\sqrt{\frac{11,49^2}{95} + \frac{10,83^2}{95}}} = 1,648$$



8) Conclusión

Siendo 1,648 > 1,645, se rechaza H$_0$. La diferencia es estadísticamente significativa, los niños de ambas poblaciones difieren en promedio en el puntaje del nivel de Ansiedad por Separación Infantil.

Tamaño del efecto

Aplicamos la fórmula del *tamaño del efecto g* de Hedges para contrastes de hipótesis sobre diferencia de medias:

$$g = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{(n_1 - 1) \cdot S_1'^2 + (n_2 - 1) \cdot S_2'^2}{n_1 + n_2 - 2}}}$$

Reemplazando con los datos del ejercicio:

$$d = \frac{29{,}66 - 26{,}99}{\sqrt{\frac{(95-1) \cdot 11{,}49^2 + (95-1) \cdot 10{,}83^2}{95 + 95 - 2}}} = 0{,}24$$

d = 0,24, por lo que el tamaño del efecto es pequeño. La separación de los padres tiene efecto, aunque pequeño, sobre el nivel de ansiedad por separación de los hijos, aumentándolo.

EJERCICIO 5

Los puntajes en una prueba objetiva de rendimiento escolar siguen la distribución normal con media 15 y varianza 11,25 bajo el supuesto de responder todos los ítems al azar. Un docente toma la prueba a un curso de 36 alumnos y obtiene un puntaje medio de 16,5 puntos. ¿Puede descartar, a un nivel del 1%, la posibilidad de que sus alumnos hayan respondido al azar?

El docente desea descartar la posibilidad de que sus alumnos hayan respondido al azar en la prueba de rendimiento escolar; se trata de una prueba bilateral.

1) Variable aleatoria

X = Puntaje obtenido por cada alumno en la prueba objetiva de rendimiento escolar.

2) Supuestos mínimos necesarios para llevar a cabo el contraste

- La variable se distribuye normalmente.
- El desvío es conocido ($\sigma^2 = 11{,}25 \implies \sigma = 3{,}3541$).
- La muestra es aleatoria.

3) Hipótesis estadísticas

H_0: $\mu = 15$ El puntaje medio en la prueba objetiva de rendimiento escolar es 15, bajo el supuesto de responder todos los ítems por azar.

H_1: $\mu \neq 15$ El puntaje medio en la prueba objetiva de rendimiento escolar es diferente de 15 (no contesta por azar).

4) Nivel de significación

Nivel de significación del 1%, ∝ = 0,01.

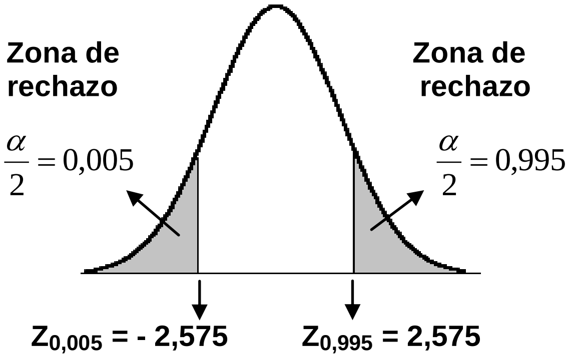
5) Estadístico de contraste y su distribución bajo H_0

$$E=\frac{\overline{X}-15}{\frac{3,3541}{\sqrt{36}}}\approx N(0,1)$$

6) Regla de decisión

Se trata de un contraste bilateral con desvío conocido. La zona de rechazo se encuentra repartida a partes iguales en ambas colas de la distribución muestral. Siendo α = 0,01, el área de la zona de rechazo en cada extremo de la distribución muestral es $\frac{\alpha}{2}=0,005$. Buscamos los dos valores críticos que dividen a la distribución en ambas zonas, zona de rechazo y zona de aceptación:

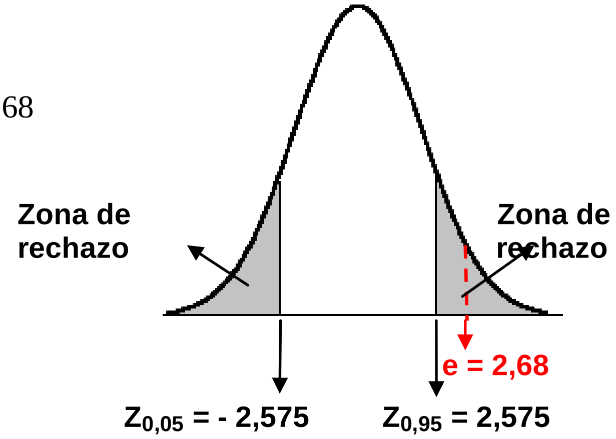
$Z_{0,005}$ = - 2,575 y **$Z_{0,995}$ = 2,575**



Se rechaza H_0 si y sólo si el valor observado del estadístico de contraste es menor que – 2,575 o mayor que 2,575.

7) Cálculo del valor observado del estadístico de contraste

$$e=\frac{16,5-15}{\frac{3,3541}{\sqrt{36}}}=2,68$$



8) Conclusión

Siendo 2,68 > 2,575, se rechaza H_0. Se descarta la posibilidad de que los alumnos hayan contestado por azar.

EJERCICIO 6

El siguiente cuadro muestra la cantidad de alumnos que han elegido cada una de las opciones de post-grado que la universidad brinda. Estos resultados permiten inferir la demanda de cursos que tendrán los alumnos que hoy cursan el último año de la carrera. Suponiendo que el año entrante tenga un comportamiento similar al anterior responda:

Especialidad	Cant
Área Clínica	150
Área Jurídica	30
Área Educacional	60
Área Comunitaria	25
Área Laboral	35
Total	300

I. Con qué probabilidad en una muestra de 7 alumnos que cursarán el posgrado en esta universidad

a) a lo sumo 5 no elijan el área Jurídica.
b) todos elijan el área Clínica.
c) ninguno elija el área Educacional.
d) la mayoría elija Laboral o Comunitaria.
e) por lo menos 2 no elijan el área Educacional.

II. ¿Puede afirmarse con un nivel de significación del 1% que para el año entrante menos de la quinta parte de los alumnos elegirán el área laboral?

I.
a) Seleccionamos una muestra aleatoria de 7 alumnos, ***n* = 7**. De acuerdo a la información brindada por el cuadro, 30 de 300 alumnos eligen cursar el posgrado en el área Jurídica. Considerando éxito la condición mencionada, la probabilidad de seleccionar un alumno que curse el posgrado en el área Jurídica es $\frac{30}{300} = 0{,}10$, ***p* = 0,1**.

Sea X = "Cantidad de alumnos que cursarán el posgrado en el área Jurídica"

$$X \sim B(7;\ 0{,}1)$$

Hallar la probabilidad de que a lo sumo 5 alumnos no elijan el área Jurídica es equivalente a hallar la probabilidad de que 2 o más alumnos sí elijan dicha área.

$$P(X \geq 2) = 1 - [P(X = 0) + P(X = 1)] = 1 - [0{,}478 + 0{,}372] = 1 - 0{,}850 = \underline{\mathbf{0{,}150}}$$

b) Se trata, ahora, de hallar la probabilidad de que todos los alumnos de la muestra elijan el área Clínica; 150 de 300 alumnos eligen cursar el posgrado en el área Clínica. Por lo tanto, la probabilidad de seleccionar un alumno que elija dicha área es $\frac{150}{300} = 0{,}50$, ***p* = 0,5**.

X = "Cantidad de alumnos que cursarán el posgrado en el área Clínica"

$$X \sim B(7;\ 0{,}5)$$

P (X = 7) = $\underline{\mathbf{0{,}008}}$

c) Sea X = "Cantidad de alumnos que cursarán el posgrado en el área Educacional"

60 de 300 alumnos eligen cursar el posgrado en el área Educacional, la probabilidad de seleccionar un alumno que elija dicha área es $\frac{60}{300} = 0{,}2$, ***p* = 0,2**.

X ~ B(7; 0,2)

P (X = 0) = **<u>0,210</u>**

d) X = "Cantidad de alumnos que cursarán el posgrado en el área Laboral o Comunitaria"

35 alumnos eligen el área Laboral y 25 el área Comunitaria. La probabilidad de seleccionar un alumno que elija cualquiera de las dos áreas es $\frac{60}{300} = 0{,}2$ (35 + 25 = 60), ***p* = 0,2**.

X ~ B(7; 0,2)

Hallar la probabilidad de que la mayoría elija alguna de estas dos áreas es equivalente a hallar la probabilidad de que lo haga más de la mitad:

$$P(X \geq 4) = P(X = 4) + P(X = 5) + P(X = 6) + P(X = 7) + P(X = 8)$$

= 0,029 + 0,004 + 0 + 0 = **<u>0,033</u>**

e) Considerando lo definido en el punto c), por lo menos dos alumnos no elijan el

área Educacional es equivalente a que sí elijan dicha área un máximo de cinco alumnos:

$$P(X \leq 5) = 1 - [P(X = 6) + P(X = 7)] = 1 - [0 +] = 1$$

Recordemos que la tabla de la binomial nos proporciona las probabilidades calculadas y redondeadas a la tercera cifra decimal. En este caso, las probabilidades correspondientes a los valores 6 y 7 de la variable están redondeados a 0. Si hubiéramos sumado las probabilidades desde X = 0 hasta X = 5, también habríamos obtenido 1 por los redondeos efectuados en la tabla.

Sin embargo, utilizando el programa Excel, el error que se comete por redondeo es menor ya que las probabilidades están calculadas y expresadas con más cifras decimales. Los resultados que hubiéramos obtenido son los que ofrece esta guía como respuesta:

I. a) 0,1497 b) 0,0078 c) 0,2097 d) 0,0333 e) 0,9996

II.

Debemos probar si, para el año entrante, menos de la quinta parte de los alumnos, es decir, menos del 20%, elegirán el área laboral entre las opciones de post-grado que brinda la universidad. Según la información que brinda el cuadro, 35 alumnos de un total de 300, eligen el área laboral, por lo tanto la proporción de éxitos en la muestra es $0{,}11\overset{\frown}{6}$.

1) Variables y supuestos mínimos necesarios para llevar a cabo el contraste

X: Condición de elegir o no el área laboral entre las opciones de post-grado, en la población de alumnos de la Facultad de Psicología de la UBA.

Supuestos: Las observaciones muestrales son independientes una de otra y con la misma probabilidad de elegir el área laboral. Es decir, para cada elección al azar, X es una variable Bernoulli con el mismo parámetro π sobre el cual se desea hacer inferencias.

2) Hipótesis estadísticas

$H_0 : \pi \geq 0{,}20$ Por lo menos la quinta parte de los alumnos en cuestión elige el área laboral.

$H_1 : \pi < 0{,}20$ Menos de la quinta parte de los alumnos en cuestión elige el área laboral.

El procedimiento es el mismo que si las hipótesis fueran:

$$H_0 : \pi = 0{,}20$$
$$H_1 : \pi < 0{,}20$$

3) Nivel de significación

Nivel de significación del 1%, ∝ = 0,01.

4) Estadístico de contraste y su distribución bajo H_0

$$E = \frac{p - 0{,}20}{\sqrt{\dfrac{0{,}20 \cdot (1 - 0{,}20)}{300}}} \approx N(0{,}1)\text{ , bajo } H_0$$

El estadístico E se distribuye normalmente por el Teorema Central del Límite ya que la muestra es grande (300 alumnos). Además se cumple que $n \cdot \pi_0 = 300 \cdot 0{,}20 = 60 > 5$ y $n \cdot (1 - \pi_0) = 300 \cdot (1 - 0{,}20) = 240 > 5$

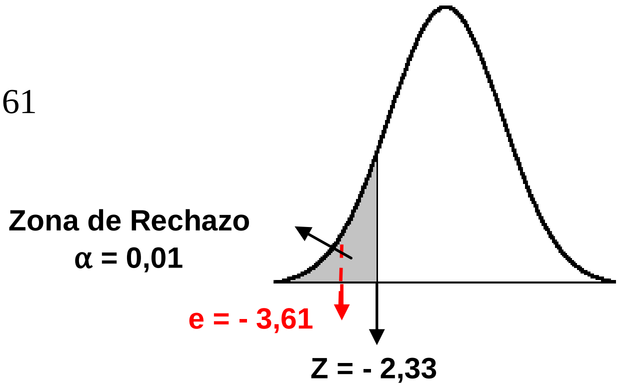
5) Regla de decisión

Es un contraste unilateral izquierdo, se rechaza H_0 si y sólo si el valor observado del estadístico de contraste es menor que $Z_{0,01}$.

Buscamos en la tabla: $Z_{0,01} = -2{,}33$

6) Cálculo del valor observado del estadístico de contraste

$$e = \frac{0{,}11\overset{\frown}{6} - 0{,}20}{\sqrt{\dfrac{0{,}20 \cdot (1 - 0{,}20)}{300}}} = -3{,}61$$



7) Conclusión

Como – 3,61 < - 2,33, se rechaza H_0; menos de la quinta parte de los alumnos elegirán el área laboral

EJERCICIO 7

Remitiéndose a los datos del ejercicio 6 de la Práctica VII pruebe, al 5%, que los CI de la población correspondiente al grupo experimental aumentaron más, en promedio, que los CI de la del grupo control, según se sugiere en el comentario de la respuesta.

Se sugiere en el comentario de la respuesta de dicho ejercicio, siempre que sea posible, tratar todos los datos del experimento conjuntamente a fin de que la decisión final obedezca a un único nivel de significación. Esto es, debemos plantear que la diferencia entre los CI del post-test y pre-test es, en promedio, mayor para la población hipotética representada por el grupo experimental que para la representada por el grupo control.

En el ejercicio 6 de la Práctica VII hemos calculado la media y el cuasidesvío de los puntajes diferencia entre los CI del post-test y pre-test, para cada grupo.

Los valores obtenidos se muestran en la siguiente tabla:

	n	$\bar{X}$	*S'*
Grupo Experimental	15	31	13,9745
Grupo Control	15	9,8	21,2677

Debemos responder al 5% si la diferencia entre los CI del post-test y pre-test es, en promedio, mayor para la población hipotética representada por el grupo experimental que para la representada por el grupo control. Se trata de una prueba unilateral a derecha.

1) Variables aleatorias

X_1: Puntaje diferencia en el CI del post-test y pre-test de cada niño de 2 a 4 años que reciben el programa de estimulación temprana (PET).

X_2: Puntaje diferencia en el CI del post-test y pre-test de cada niño de 2 a 4 años que no reciben el programa de estimulación temprana (PET).

2) Supuestos

- Los puntajes, X_1 y X_2, se distribuyen normalmente.
- Se extraen dos muestras aleatorias e independientes de las dos poblaciones, cada una de 15 sujetos.
- Las varianzas son desconocidas e iguales.

3) Hipótesis estadísticas

H_0: $\mu_{DE} = \mu_{DC}$ En ambas poblaciones la diferencia media de CI del pre al post test es igual.

H_1: $\mu_{DE} > \mu_{DC}$ El aumento, en promedio, del CI de quienes reciben el PET es mayor que el de quienes no lo reciben.

o, equivalentemente,

H_0: $\mu_{DE} - \mu_{DC} = 0$
H_1: $\mu_{DE} - \mu_{DC} > 0$

donde μ_{DE} es la media de la diferencia para la población representada por el grupo experimental y μ_{DC} la correspondiente para el grupo control.

4) Nivel de significación

Nivel de significación del 5%, ∝ = 0,05.

5) Estadístico de contraste y su distribución bajo H_0

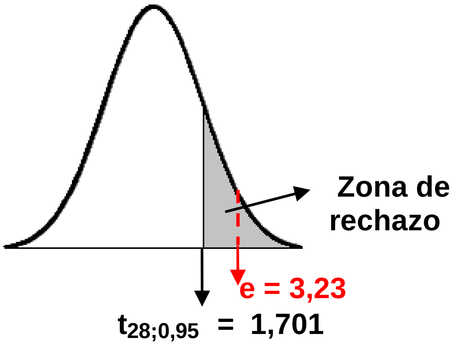
$$E = \frac{\overline{X}_{DE} - \overline{X}_{DC}}{\sqrt{\frac{14 \cdot S'^2_1 + 14 \cdot S'^2_2}{15+15-2} \cdot \left(\frac{1}{15} + \frac{1}{15}\right)}} \sim t_{28} \text{ bajo } H_0 \text{ (gl = 15 + 15 – 2)}$$

6) Regla de decisión

Es un contraste unilateral derecho, se rechaza H_0 si y sólo si el valor observado del estadístico de contraste es mayor que $t_{28;0,95}$.

7) Cálculo del valor observado del estadístico de contraste

$$e = \frac{31 - 9,8}{\sqrt{\frac{14 \cdot 13,9745^2 + 14 \cdot 21,2677^2}{15+15-2} \cdot \left(\frac{1}{15} + \frac{1}{15}\right)}} = 3,23$$



8) Conclusión

Siendo **3,23** > 1,701 se rechaza H_0. El programa es eficaz porque los niños que lo recibieran aumentarían su CI más que aquellos que sólo tuvieran la estimulación ordinaria.

EJERCICIO 8. Artículo completo en http://www.psicothema.com/pdf/3523.pdf

Rosario et al (2008) estudian la ansiedad en los exámenes en relación al sexo, entre otras variables personales y familiares. La ansiedad se evaluó con el Cuestionario de Ansiedad ante los Exámenes (CAE), de Rosario y Soares (2004), el cual está destinado a los alumnos entre 11 y 15 años que cursan la Enseñanza Secundaria Obligatoria (ESO). El cuestionario fue administrado a una muestra compuesta por 258 mujeres y 275 varones estudiantes de la ESO. Los autores exhiben la siguiente tabla:

Sexo	Media	Desvío típico	n
Varones	24,26	7,14	275
Mujeres	27,02	6,96	258

¿Puede afirmarse, con un nivel de significación del 5%, que hay diferencias estadísticamente significativas entre mujeres y varones? Si es así, calcule el tamaño del efecto y concluya si tal diferencia es psicológicamente significativa.

Debemos responder al 10% si la diferencia observada en el nivel de ansiedad por separación infantil es estadísticamente significativa. Se trata de una prueba bilateral.

1) Variables aleatorias

X_1: Puntaje de cada estudiante mujer de la Enseñanza Secundaria Obligatoria en el Cuestionario de Ansiedad ante los Exámenes.

X_2: Puntaje de cada estudiante varón de la Enseñanza Secundaria Obligatoria en el Cuestionario de Ansiedad ante los Exámenes.

2) Supuestos

Las muestras seleccionadas son grandes, no es necesario suponer la normalidad de los puntajes en cada población ni la homogeneidad de las varianzas.

3) Hipótesis estadísticas

$H_0 : \mu_1 - \mu_2 = 0$ Los estudiantes de ambas poblaciones no difieren en promedio en el puntaje del nivel de Ansiedad ante los exámenes.

$H_0 : \mu_1 - \mu_2 \neq 0$ Los estudiantes de ambas poblaciones difieren en promedio en el puntaje del nivel de Ansiedad ante los exámenes.

4) Nivel de significación

Nivel de significación del 5%, $\propto = 0,05$.

5) Estadístico de contraste y su distribución bajo H_0

$$E = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^2}{258} + \frac{S_2^2}{275}}} \approx N(0,1)$$

por un teorema similar al Teorema Central del Límite ya que cada muestra es grande.

Bajo H_0

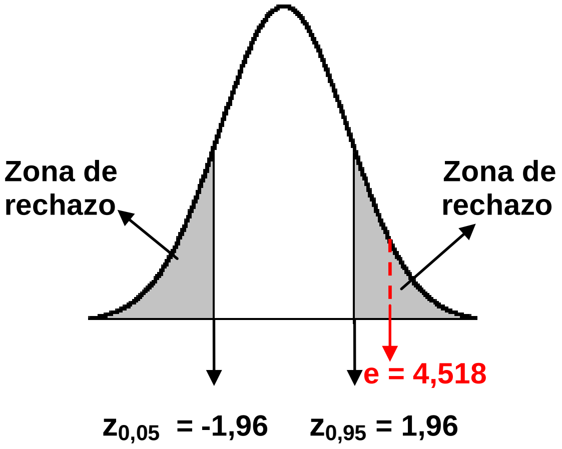
6) Regla de decisión

Es un contraste bilateral, se rechaza H_0 si y sólo si el valor observado del estadístico de contraste es menor que $z_{0,025}$ o mayor que $z_{0,975}$.

Buscamos en la tabla de la normal los valores de z: $z_{0,025} = -1,96$, $z_{0,975} = 1,96$.

7) Cálculo del valor observado del estadístico de contraste

$$e = \frac{27,02 - 24,26}{\sqrt{\frac{6,96^2}{258} + \frac{7,14^2}{275}}} = 4,518$$



8) Conclusión

Siendo 4,518 > 1,96, se rechaza H_0, por lo que la diferencia es estadísticamente significativa.

Tamaño del efecto

$$d = \frac{27,02 - 24,26}{\sqrt{\frac{(258-1)\cdot 6,96^2 + (275-1)\cdot 7,14^2}{258 + 275 - 2}}} = 0,39$$

El tamaño del efecto es 0,39, apenas moderado, siendo las mujeres las que presentan un nivel de ansiedad en promedio algo mayor.

EJERCICIO 9

Gastón se recibió recientemente de psicólogo y va a abrir un consultorio pero no se siente muy seguro en la atención de casos con síndrome de depresión. Consultando entre sus colegas obtuvo la siguiente información:

	Consulta A	Consulta B	Consulta C	Consulta D
Casos síndrome de depresión	3	5	15	7
Total de pacientes	30	55	135	80

En el hipotético caso de que al consultorio de Gastón acudieran 9 pacientes ¿qué probabilidad le asignaría al hecho de que

a) alguno tenga síndrome de depresión?
b) más de la mitad no tenga síndrome de depresión?

a) A los cuatro consultorios asisten un total de 300 pacientes y de ellos presentan síndrome de depresión 30 pacientes.

Sea X = “Cantidad de pacientes con síndrome de depresión”

Llamando éxito a la condición de presentar síndrome de depresión, la probabilidad de éxito es $\frac{30}{300} = 0{,}1$; ***p* = 0,1**

La muestra está formada por los 9 pacientes que acuden al consultorio de Gastón, por lo tanto, ***n* = 9**.

X ~ B(9; 0,1)

Debemos hallar la probabilidad de que alguno de los pacientes tenga síndrome de depresión, esto es, que por lo menos uno de ellos presente este síndrome.

$P(X \geq 1) = 1 - [P(X = 0)] = 1 - 0{,}387 =$ **<u>0,613</u>**

b) Que más de la mitad no tenga síndrome de depresión es equivalente a que la cantidad de sujetos que no tenga síndrome de depresión sea mayor o igual a 5, y esto es equivalente a que sí presenten este síntoma 4 o menos sujetos.

$P(X \leq 4) = 1 - P(X \geq 5) = 1 - 0{,}001 =$ **<u>0,999</u>**

EJERCICIO 10

En una investigación donde se analizan los efectos del video juego *Tradislexia* en la conciencia fonológica y reconocimiento de palabras en niños disléxicos, la hipótesis es que el entrenamiento multimedia contribuye a mejorar los procesos fonológicos. Para poner a prueba esta hipótesis se seleccionaron 32 niños con problemas de dislexia entre 9 y 12 años del segundo y tercer ciclo de educación primaria, y se asignaron al azar a un grupo experimental y a un grupo control, de igual tamaño (16). Al grupo experimental se lo entrenó con el video juego Tradislexia, 30 minutos diarios, en un total de 20 sesiones mientras que el grupo control recibía la instrucción común de la escuela. Al cabo del período de entrenamiento se les administró la subescala de lectura de palabras (correcta identificación de 30 palabras) de la batería de Evaluación de los Procesos Lectores de los Niños de Educación Primaria PROLEC (Cuetos et al., 1996). Se obtuvieron los siguientes datos:

Grupo Experimental:
21 – 15 – 18 – 13 – 17 – 21 – 20 – 14 – 18 – 16 – 24 – 10 – 23 – 11 – 14 – 17
Grupo Control:
11 – 8 – 5 – 17 – 3 – 19 – 10 – 13 – 8 – 14 – 15 – 11 – 7 – 12 – 9 – 14

a) Enuncie la hipótesis científica y su formulación en términos estadísticos.

b) Lleve a cabo el contraste de hipótesis correspondiente con un nivel de significación del 10% y exprese una conclusión. Calcule el tamaño del efecto, si lo hay.

c) ¿Qué supuestos sobre la distribución de los puntajes en las correspondientes poblaciones son necesarios para validar el contraste?

Nota: *Este ejercicio es una adaptación, con datos ficticios, de la investigación de* Jiménez, J. y Rojas, E. (2008). El artículo completo puede consultarse en http://www.psicothema.com/pdf/3491.pdf

a) *Hipótesis científica*: el entrenamiento multimedia contribuye a mejorar los procesos fonológicos.

Hipótesis estadística: $\mu_1 - \mu_2 > 0$

donde μ_1 es la media en la subescala PROLEC de los niños disléxicos españoles si recibieran el entrenamiento multimedia para el grupo experimental, y μ_2 es la media en la subescala PROLEC de los niños disléxicos españoles si sólo recibieran el entrenamiento escolar (grupo control).

b) y c) Debemos probar, con un nivel de significación del 10%, si el entrenamiento multimedia contribuye a mejorar los procesos fonológicos; se trata, pues, de una prueba unilateral a derecha.

Hallamos los resúmenes estadísticos correspondientes a cada grupo:

<u>Grupo Experimental</u>:

X_i	n_i	$X_i \cdot n_i$	$\left(x_i - \bar{x}\right)^2 \cdot n_i$
10	1	10	49
11	1	11	36
13	1	13	16
14	2	28	18
15	1	15	4
16	1	16	1
17	2	34	0
18	2	36	2
20	1	20	9
21	2	42	32
23	1	23	36
24	1	24	49
Total	**16**	**272**	**252**

Media aritmética: $\bar{x} = \frac{272}{16} = 17$

Cuasidesvío: $S' = \sqrt{\frac{252}{15}} = 4{,}0988$

Grupo de Control:

X_i	**n_i**	**$X_i \cdot n_i$**	$\left(x_i - \overline{x}\right)^2 \cdot n_i$
3	1	3	64
5	1	5	36
7	1	7	16
8	2	16	18
9	1	9	4
10	1	10	1
11	2	22	0
12	1	12	1
13	1	13	4
14	2	28	18
15	1	15	16
17	1	17	36
19	1	19	64
Total	**16**	**176**	**278**

Media aritmética: $\overline{x} = \frac{176}{16} = 11$

Cuasidesvío: $S' = \sqrt{\frac{278}{15}} = 4{,}305$

Resumiendo los cálculos hechos:

	n	$\overline{X}$	*S'*
Grupo Experimental	16	17	4,0988
Grupo de Control	16	11	4,305

1) Variables aleatorias

X_1: Puntaje en la subescala PROLEC de los niños disléxicos españoles si recibieran el entrenamiento multimedia.

X_2: Puntaje en la subescala PROLEC de los niños disléxicos españoles si sólo recibieran el entrenamiento escolar.

2) Supuestos

X_1 y X_2 son variables aleatorias independientes, normalmente distribuidas y con igual varianza.

3) Hipótesis estadísticas

$H_0 : \mu_1 - \mu_2 = 0$ Los sujetos de ambas poblaciones no difieren en promedio en su puntaje en la subescala PROLEC.

$H_0 : \mu_1 - \mu_2 > 0$ Los sujetos de ambas poblaciones difieren en promedio en su puntaje en la subescala PROLEC.

4) Nivel de significación

Nivel de significación del 10%, $\propto = 0{,}10$.

5) Estadístico de contraste y su distribución bajo H₀

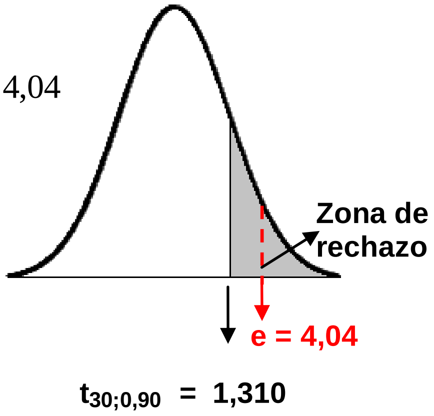
$$E = \frac{\overline{X}_{EJ} - \overline{X}_{EM} - 2}{\sqrt{\frac{15 \cdot S'^2_1 + 15 \cdot S'^2_2}{16+16-2} \cdot \left(\frac{1}{16} + \frac{1}{16}\right)}} \quad \sim \quad t_{30} \text{ bajo } H_0 \ (gl = 16 + 16 - 2)$$

6) Regla de decisión

Es un contraste unilateral derecho, se rechaza H_0 si y sólo si el valor observado del estadístico de contraste es mayor que $t_{30;0,90}$.

7) Cálculo del valor observado del estadístico de contraste

$$e = \frac{17 - 11}{\sqrt{\frac{(16-1) \cdot 4{,}0988^2 + (16-1) \cdot 4{,}305^2}{16+16-2} \cdot \left(\frac{1}{16} + \frac{1}{16}\right)}} = 4{,}04$$



8) Conclusión

Siendo 4,04 > 1,310, se rechaza H_0. El entrenamiento multimedia contribuye a mejorar los procesos fonológicos.

Tamaño del efecto

$$d = \frac{17 - 11}{\sqrt{\frac{(16-1) \cdot 4{,}0988^2 + (16-1) \cdot 4{,}305^2}{16+16-2}}} = 1{,}43$$

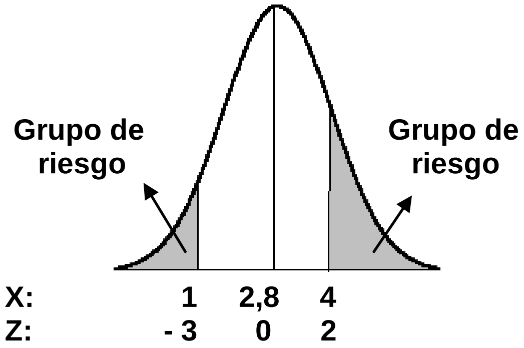
El tamaño del efecto es 1,43 que es alto.

EJERCICIO 11

Un peso inferior al kilo o superior a los cuatro kilos al nacer puede considerarse riesgoso en una población determinada. Suponiendo que los pesos se distribuyan normalmente en los recién nacidos con media 2,8 y desvío 0,6 kilogramos. Calcule la probabilidad de que un recién nacido

a) esté dentro del grupo de riesgo.
b) pese entre 2 y 3 kilogramos.

a) El grupo de riesgo está formado por los recién nacidos que tengan un peso inferior a 1 kg o superior a 4 kg. Gráficamente:



Transformamos los puntajes que limitan el grupo de riesgo en puntajes Z:

$$z = \frac{1-2,8}{0,6} = -3 \qquad\qquad z = \frac{4-2,8}{0,6} = 2$$

Buscamos en la tabla de la normal el área a izquierda de z = -3 y z = 2.

El área a izquierda de z = - 3 es 0,0013.

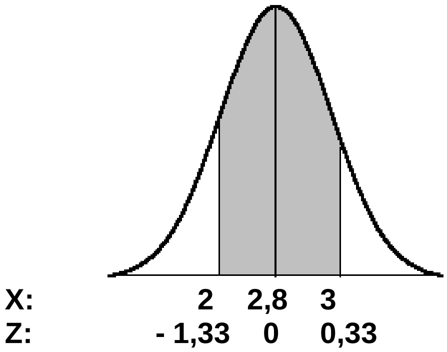
El área a izquierda de 2 es 0,9772, por lo que el área que corresponde al grupo de riesgo que tiene un puntaje mayor a 2 se obtiene restando de 1 este valor:

1 – 0,9772 = 0,0228

La suma de ambas áreas es la probabilidad de que un recién nacido pertenezca al grupo de riesgo: 0,0013 + 0,0228 = **0,0241**

b) Para hallar la probabilidad de que un recién nacido pese entre 2 y 3 kg, calculamos los puntajes z correspondientes a 2 y 3 kg:

$$z = \frac{2-2,8}{0,6} = -1,\widehat{3} \qquad\qquad z = \frac{3-2,8}{0,6} = 0,\widehat{3}$$



Buscamos en la tabla de la normal el área a izquierda de z = - 1,33 y z = 0,33.

El área a izquierda de z = - 1,33 es 0,0918.

El área a izquierda de z = 0,33 es 0,6293,

La probabilidad de que un recién nacido tenga un peso entre 2 y 3 kilogramos se obtiene restando ambas áreas:

0,6293 - 0,0918 = **0,5375**

EJERCICIO 12

Una prueba de oído musical consiste en reproducir el sonido de la nota “la” (que corresponde a una frecuencia de 440) luego de oírla. Un grupo de 10 personas elegidas aleatoria e independientemente entre profesionales que se dedican a la afinación de instrumentos musicales realizó la prueba y de las 10 emisiones se obtuvo una frecuencia media de 435 con un desvío estándar de 11. ¿Puede afirmarse que la frecuencia media de las notas que emitirían los profesionales de la población muestreada difiere de 440 al nivel de 10%? Suponga la distribución normal de las frecuencias emitidas.

Disponemos de los siguientes datos:

	DATOS DE LA POBLACIÓN	DATOS DE LA MUESTRA
Desvío desconocido	μ = 440	$\bar{x} = 435$ n = 10 S' = 11

1) Variable aleatoria

X = Frecuencia de reproducción, emitida por cada profesional que se dedica a la afinación de instrumentos musicales, del sonido de la nota “la” en la prueba de oído musical.

2) Supuestos mínimos necesarios para llevar a cabo el contraste

- La variable se distribuye normalmente.
- El desvío es desconocido.
- La muestra es aleatoria.

3) Hipótesis estadísticas

H_0: μ = 440 La frecuencia media de las notas emitidas por los profesionales es igual a 440.

H_1: μ ≠ 440 La frecuencia media de las notas emitidas por los profesionales es diferente de 440.

4) Nivel de significación

Nivel de significación del 10%, ∝ = 0,10.

5) Estadístico de contraste y su distribución bajo H_0

$$E = \frac{\overline{X} - 440}{\frac{11}{\sqrt{10}}} \quad \sim t_9 \text{ bajo } H_0 \quad (gl = 10 - 1 = 9)$$

6) Determinar la zona de rechazo de H_0 o “regla de decisión”

Se trata de un contraste bilateral con desvío desconocido. El nivel de significación es $\alpha = 0{,}10$, por lo que el área de la zona de rechazo en cada extremo de la distribución muestral es $\frac{\alpha}{2} = 0{,}05$.

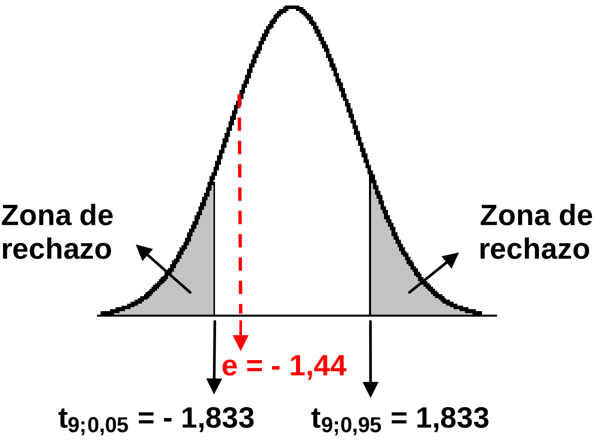
Buscamos en la tabla t de *Student*, **$t_{9;0,05}$ = - 1,833** y **$t_{9;0,95}$ = 1,833**

7) Regla de decisión

Se rechaza H_0 si y sólo si el valor observado del estadístico de contraste es menor que - 1,833 o mayor que 1,833.

8) Cálculo del valor observado del estadístico de contraste

$$e = \frac{435 - 440}{\frac{11}{\sqrt{10}}} \approx -1{,}44$$



9) Conclusión

Siendo - 1,44 > - 1,833, no se rechaza H_0. La diferencia observada puede atribuirse a razones fortuitas y no necesariamente a que los profesionales emitan en promedio una frecuencia diferente de la correcta para la nota “la”.

EJERCICIO 13. Artículo completo en http://www.psicothema.com/pdf/3593.pdf

Errasti Pérez et al (2009) estudian la eficacia de un programa para la prevención del consumo de drogas. Lo pusieron a prueba sobre 378 adolescentes de escuelas secundarias de Asturias a los que administraron, entre otras escalas, un cuestionario de consumo de drogas antes y después de la implementación del programa. Obtuvieron una diferencia media entre los puntajes del post-test menos los del pre-test de -0,4067 con una desviación estándar de 2,8961. ¿Puede afirmarse con un nivel de significación del 1% que el programa es eficaz? En caso afirmativo dé una medida del efecto.

Se desea estudiar la eficacia de un programa para la prevención del consumo de drogas, para lo cual, se aplica un cuestionario de consumo de drogas antes y después de la implementación del programa sobre una muestra de 378 adolescentes de escuelas secundarias de Asturias. Se trata, pues, de una prueba unilateral a izquierda para diferencia de medias de datos apareados.

1) Variable aleatoria

D = Diferencia entre los puntajes del cuestionario de consumo de drogas, de cada adolescente de escuelas secundarias de Asturias, antes y después de la implementación del programa.

2) Supuestos

- la variable $D = X_D - X_A$ se distribuye normalmente con media μ_D y desvío estándar σ_D desconocido.
- la muestra es aleatoria.

3) Hipótesis estadísticas

$H_0 : \mu_D = 0$ El programa para la prevención del consumo de drogas no es eficaz.

$H_0 : \mu_D < 0$ El programa para la prevención del consumo de drogas es eficaz

μ_D denota la media de las diferencias entre los puntajes después y antes de la implementación del programa para la prevención del consumo de drogas.

4) Nivel de significación

Nivel de significación del 1%, ∝ = 0,01.

5) Estadístico de contraste y su distribución bajo H_0

por el Teorema Central del Límite ya que la muestra es grande.

$$E = \frac{\overline{D}}{\frac{S'}{\sqrt{378}}} \approx N(0,1)$$

Bajo H_0

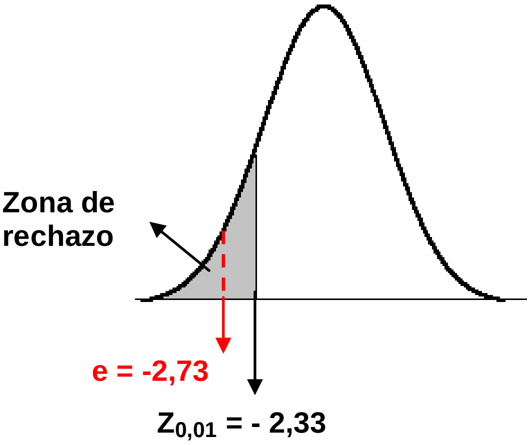
6) Regla de decisión

Es un contraste unilateral izquierdo, se rechaza H_0 si y sólo si el valor observado del estadístico de contraste es menor que $z_{0,01}$.

Buscamos en la tabla: $z_{0,01} = - 2,33$

7) Cálculo del valor observado del estadístico de contraste

$$E = \frac{-0{,}4067}{\frac{2{,}8961}{\sqrt{378}}} = -2{,}73$$



8) Conclusión

Siendo - 2,73 < - 2,33, se rechaza H_0.

Tamaño del efecto

$$d = \frac{-0{,}4067}{2{,}8961} \approx 0{,}14$$

El tamaño del efecto es 0,14, pequeño. El programa tiene eficacia pero con un efecto pequeño.

REFERENCIAS

Cuetos, F., Rodríguez, B., y Ruano, E. (1996). *Batería de evaluación de los procesos lectores de los niños de Educación Primaria* (PROLEC). Madrid: T.E.A., Ediciones.

Errasti Pérez, J.; Al-Halabí Díaz, S.; Secades Villa, R.; Fernández-Hermida, J.; Carballo, J. y García-Rodríguez, O. (2009). Prevención familiar del consumo de drogas: el programa «Familias que funcionan. *Psicothema, 21*, 1, 45-50.

Espada, J., Méndez, F., Orgilés, M. e Hidalgo, M. (2006). Cuestionario de ansiedad por separación infantil, Forma Niños. Documento policopiado. Alicante: Universidad Miguel Hernández.

Jiménez, J. y Rojas, E. (2008). Efectos del videojuego Tradislexia en la conciencia fonológica y reconocimiento de palabras en niños disléxicos. *Psicothema, 20*, 3, 347 – 353.

Orgilés Amorós, M., Espada Sánchez, J. y Méndez Carrillo, X. (2008). Trastorno de ansiedad por separación en hijos de padres divorciados. *Psicothema, 20*, 3, 383-388.

Rosario, P., y Soares, S. (2004). Questionário de Ansiedade face aos Testes (QAT). En Leandro, A., Simões, M., Machado, C., y Gonçalves, M. (Eds.): *Avaliação Psicológica: Instrumentos validados para a população portuguesa* (vol. II, pp. 39-51). Coímbra: Quarteto Editora.

Rosario, P.; Núñez, J.; Salgado, A.; González-Pienda, J.; Valle, A.; Joly, C. y Bernardo, A. (2008). Ansiedad ante los exámenes: relación con variables personales y familiares. *Psicothema*, 20, 4, 563-570.